VISIT…

LANZAROTE Caliente.COM

《语言学论丛》编辑委员会

（按姓氏笔画排列）

王　力（主任委员）　王福堂　石安石
朱德熙　陆俭明　岑麒祥　林　焘
周祖谟　唐作藩　曹先擢　蒋绍愚

主　　编：林　焘
责任编辑：王福堂

编者按: 美国加州大学(伯克利)王士元 (William S-Y. Wang) 教授,曾应北京大学邀请,于 1979 年夏到北京大学讲学。在五周时间内,王士元教授系统讲授了“实验语音学”课程,并做了“语言的发生”等四次学术报告。在此以后,北大中文系语音实验室根据录音将课程和报告整理成文,印发了少量打印本。整理工作是由林焘、王福堂、王理嘉三位同志负责的,几位研究生和实验室工作人员参加了录音记录工作。“实验语音学”原来共讲十二课,整理稿归纳为八讲和两个附录,并根据内容加上了标题。最近王士元教授审阅了这份整理稿,并同意在《语言学论丛》第十一辑上作为专辑发表。对此,我们谨向作者表示感谢。我们希望,本书将能向读者比较全面地介绍实验语音学等方面的基本知识,其中不少事实和观点并能对人有所启发。

目　　录

实验语音学讲座 ……………………………〔美〕王士元（3）
　一　语言学和语音学 ……………………………（3）
　二　发音过程和发音器官 ……………………………（7）
　三　语音的物理分析 ……………………………（13）
　四　语音中的声压和共振峰 ……………………………（23）
　五　语音的基频、共振峰和元音的关系 ……………………………（34）
　六　语音的外来变化和内在变化 ……………………………（44）
　七　语音的图谱 ……………………………（53）
　八　听觉 ……………………………（70）
　〔附录一〕　关于区别性特征理论 ……………………………（86）
　〔附录二〕　关于声调语言 ……………………………（98）
学术报告 ……………………………〔美〕王士元（104）
　一　语言的发生 ……………………………（104）
　二　语言的演变 ……………………………（115）
　三　语言和文字的生理基础 ……………………………（130）
　四　近几十年来的美国语言学 ……………………………（145）

CONTENT

Lectures on Experimental Phonology

……………………………………… *William S-Y. Wang* (3)

1. Linguistics and phonetics …………………………………… (3)
2. Articulation and organs of speech ………………………… (7)
3. Physics of speech sound …………………………………… (13)
4. Sound pressure and formant………………………………… (23)
5. Fundamentals, formants and vowels ……………………… (34)
6. Extrinsic variation and intrinsic variation………………… (44)
7. Spectrogram ………………………………………………… (53)
8. Perception of speech ……………………………………… (70)

Appendix I: Distinctive feature theory………………………… (86)

Appendix II: Tone language …………………………………… (98)

Report on Recent Researches ………… *William S-Y. Wang* (104)

1. The emergence of language ………………………………… (104)
2. Language change …………………………………………… (115)
3. Psycholinguistic studies on Chinese writing ……………… (130)
4. Recent developments in American linguistics …………… (145)

实验语音学讲座

〔美〕王 士 元

一 语言学和语音学

语言学和其他学科的关系

我们讨论的中心是实验语音学，但是我想先谈谈语言学和别的学科的关系，希望大家有个全盘的看法，不要把它们看成是互相脱节的。因为语音学到底只是语言学的一个部门。

语言是人类交流思想最重要的工具，牵涉到人类的种种活动和行为，因此要涉及许多其他学科。把知识分门别类本来就会遇到不少麻烦，研究语言的时候更是如此。我以为，大学里分别系和系不应该成为知识交流的阻碍。在这方面，美国的大学和组织有它方便的地方。美国大学有所谓"多学科" (multi-disciplinary) 的学生。这种学生学习的课程伸缩性比较大，一个学生可以同时念几个不同系的课程，学习语言学的可以和别的学科有比较多的接触机会。比如，我们伯克利加州大学语言学系的学生，可以用一半时间在系里学习，其他时间到旧金山的大医院去做语言治疗工作。不同系的教授也常常在一起合作交流。我们系有一位教授和人类学系的一位教授合作。我自己前几年也曾经和心理学系的一位教授一起教课，教多学科的课。这样，知识的分科就不会阻碍几门学科之间的知识交流了。

据我观察，语言学家和别的学科有三种不同的交流形式。一种是共同的材料，语言学家和别的专家用共同的材料。比如加州

大学，除了语言学系以外，还有英语系、俄语系、希腊语系等等，我们常常和他们交流材料，研究古俄语、古希腊语等等。第二种是共同的概念。语言学是社会科学的一部分，社会科学里有很多学科都跟语言学有同样的问题，用同样的概念。语言学跟人类学、心理学、历史学、社会学、生物学，还有哲学、新闻学等都有关系，相互之间都应该有密切的交流。需要跟谁合作，那要看研究的问题和目的是什么。比如，研究语义大概就跟动物学家和生物学家有关系。如果是研究某种语言的演变，研究它怎样受社会种种因素的影响而发生变化，研究语言的发生以及人类语言和动物“语言”的关系，那么也会和动物学家和生物学家发生关系。第三种是共同的工具，语言学家和别的学科的人使用共同的工具。比如伯克利加州大学语言学系的语音实验室里常有受过电子学训练的工程师来工作。现在电子计算机和我们实验室有极其密切的关系。有了电子计算机，可以用它做很多工作。我们实验室还有数学家和我们一起工作。我们还常常跟物理系的人，尤其是搞声学的人，在一块儿讨论问题。他们也很欢迎跟我们合作，因为语言是那么重要的一种现象。物理学家、电子学家、数学家都很自然地对语言现象有着强烈的好奇心，所以常常跟语言学家一起讨论研究这方面的问题。

由于语言学牵涉到许多别的学科，所以最近十几年来出现了这么一个名词，叫做“连接号语言学”(Hyphenated linguistics)，就是把语言学和别的学科用连接号“-”连在一起成为一门新学问。它包括的范围很广。比如“人种语言学”(Ethno-linguistics)跟人类学关系特别密切，“心理语言学”(Psycho-linguistics)、“社会语言学”(Socio-linguistics)有系统地研究在某种社会形式下语言或语音是怎样变迁的，“神经语言学”(Neuro-linguistics)研究语言和神经的关系，以及语言怎样因为神经上受到打击而发生障碍，等

等。神经语言学是一门相当新的学问，从事这方面研究的语言学家有很多都到医院里去工作。

从这里我们也可以看到美国语言学发展的背景。现在美国各大学有好几十个语言学系，其中很多是从别的学科分化出来的。也就是说，美国各大学的语言学系有很多不同的来源。除了大学的语言学系之外，其他单位也有一些语言学工作人员。有些大工业的研究所，比如 IBM (International Business Machines Corporation, 国际商用机器公司)，就有十几个语言学家在那儿做研究工作。还有贝尔电话公司 (Bell Telephone Corporation)，也有很多语言学家分成好多小组在进行研究工作。我的语音学老师 G. E. Peterson 就是贝尔电话公司出身的电子工程师。除了大工业之外，医院和教育机关也有很多语言学家在工作。语言学甚至和法律也有关系。

美国有两个比较大的跟语言学有关的组织。一个是美国语言学会，会员大概至少有五六千人，是世界上最大的语言学会组织。它除了召开学会之外，还办有学报，学报的名称是《语言》(Language)。此外，它还有一个作用，就是每年夏天办一期语言学讲习所。在语言学刚兴起的时候，许多大学都感到这方面的专家不够。当时的大学，有的研究亚洲语言的专家比较多，有的研究非洲语言的专家比较多，有的对心理学的研究比较强，但要有一个全面的语言学训练却不是很容易的事。所以，美国语言学会每年办一次语言学讲习所，每次八个星期，请世界上许多有声望的语言学家来教课和讲演，把对语言学有兴趣的学生集中在一起听讲。第一期只有几十人，现在每期至少有两三千人，发展很快。语言学讲习所每年要换一个新地方，比如去年在美国伊利诺州，前年在夏威夷，大前年在麻省。

除了美国语言学会之外，还有一个规模相当大的语言学组织，

叫做“美国语言学中心” (CAL)。这个中心完全是私人组织，主要作用是在行政和组织方面。语言学家有这么多，他们之间怎样互相联系，语言学这门学科跟政治、社会等等的关系应该是怎样的，过去十几年，“美国语言学中心”就是管这些事情的。

通过上面的介绍，我们可以看到语言学和其他学科有着广泛的联系，对社会是有种种贡献的。语言学在哪些方面能够做出贡献，就要看我们所研究的是语言学的哪个部分。

语音学的三个部分

言语的传递过程大致可以分为三个阶段：发音——声波——听觉。首先是发音。这是一种生理活动。其次是声波，发音器官动作的结果造成了空气振动，也就产生了振动的波。这是一种物理现象。然后是听觉，声波在空气中传播，到达听话人的耳朵，通过神经系统传达到大脑，形成了听觉。这也是一种生理心理活动。

根据言语传递的三个不同的阶段，语音学也分成三个不同的部分：一，发音语音学 (Articulatory phonetics)，二，声学语音学 (Acoustic phonetics)，三，听觉语音学或心理语言学 (Auditory phonetics or Psycholinguistics)。发音语音学就是传统的语音学，它从语音的生理方面进行研究，主要通过观察发音器官，依靠某些仪器的帮助，把每个声音的发音部位和发音方法确定下来。声学语音学是四十年代才开始系统地建立起来的新学科，它研究语音的物理性质以及它跟发音器官的关系。这方面的研究由于第二次世界大战后声谱仪 (spectrograph) 等电子声学仪器的发明而大有进展。由于有了这些仪器，我们对语音的声学本质了解得更深了。这方面成果的进一步运用，就促进了声音的模拟、言语的合成、计算机识别语音等方面的研究。心理语言学是最新的一门学科，它以言语传递的起点和终点——大脑作为研究对象。主要是要了解

在大脑控制下产生和处理语音的步骤和方式，以及语言信息在大脑里储存的部位和形式。声音是通过耳朵传达到大脑的，因此也要了解人耳的构造以及它在传递声波时的一些特性，这样才能更好地了解听觉系统。

发音、声波、听觉这三者之间的关系是很复杂的。自然，有时候也确实很简单，只是一对一的关系。那就是发音器官特定的形状和方法形成了特定的声波，传达到大脑成为特定的听觉。但是它们更多的时候不是一对一的关系。比如，平时我们发音，嘴里没有东西，发 [i] 就是 [i]，发 [u] 就是 [u]，可以区别得很清楚。如果在嘴里衔支铅笔，还是可以发音，可以说话，而且也能听清楚。既然嘴里衔着铅笔，咬紧了，下巴不能动了，自然和嘴里没有东西的时候发音动作不一样。但是，我们还是照样可以发出那一类音，听的人也不会听错。在这种情况下，发音、声波、听觉三者之间就不是一对一的关系，也许是二对一，或者是三对一的关系。实验证明，不同的发音可以产生同样的声波，而不同的声波在听觉上有时又可以象同一个音。总之，传统语音学对语音的描述现在看来似乎是太简单了。说话是人类行为动作中最复杂最奥妙的现象。发音、声波、听觉三者的关系很复杂，现在还没有弄清楚。它所涉及的问题不是语言学一个学科所能解决的，还有待于许多学科的专家共同努力。

二　发音过程和发音器官

发音过程和反馈的作用

用语言进行交际的过程可以简单地用图 2-1 来表示:

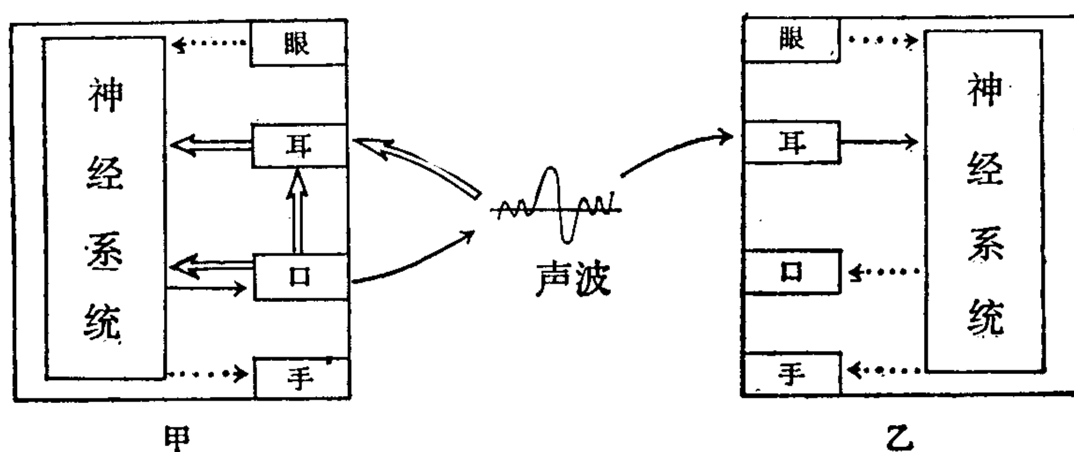


图 2-1

甲向乙表达一个意思，他需要通过大脑的神经系统发出一个指令，让发音器官做各种不同的动作，发出声音。声音经过空气的传播，送到乙的耳朵里，使乙的耳膜发生跟声波同样的振动，并且传达到乙的大脑的神经系统。通过分析，乙把意思确定了下来。然后，乙的神经系统又通知自己的发音器官，命令它做什么样的动作，发出声波，和甲进行会话。这是人类特有的一种能力。用语音交流思想只是人类交流思想最主要的一种手段。人类的文明已经有好几千年的历史，思想也可以不用语音，而用文字、光波或手势来表示。在这种情形下，就不是用嘴，而是用手把语言表达出来，通过眼睛，也一样可以传达到神经系统。

发音的时候，声波不仅和听话人有关系，而且和说话人自己也有关系。这后一种关系叫反馈 (feedback)。一般说来，反馈有两种：一种是声音上的反馈，也就是声波发出以后对自己的听音器官有反馈。这是通过两条路线实现的。一条是体外的路线，从自己的耳道里传进音去。一条是体内的，肌肉、骨头和皮肤也可以传音。通过这两条路线，可以知道自己刚才说了些什么。另一种是肌肉上的反馈，是“动觉的” (kinaesthetic)。说话的时候，嘴、舌头、软腭都在动。每个器官都有它的肌肉，肌肉的纤维一方面牵动

器官，一方面告诉自己的神经系统哪个器官在做什么动作。这种反馈完全是生理上的。

反馈的重要性已经通过语音实验得到了充分的证明。这种实验叫做延迟听觉反馈实验(delayed auditory feedback experiments)。(见图 2-2)在实验的时候，控制录音机的放音磁头，使声波的体外这条路线的反馈延迟半秒钟到达发音人自己的耳朵里，这样，在大脑里就会引起冲突。比如，声波的内部反馈路线以及生理上的反馈都告诉大脑已经说到第三个音段，可是声波的外部反馈，因为故意加以延迟，告诉大脑只说到了第一个音段。在这种情况下，说话人通常就会口吃，觉得说话很困难，有的人甚至说不出话来，因为反馈的关系被破坏了。当然，人和人之间在神经功能上并不是完全相同的，也有少数人，比如五个人当中有一个，他们说话并不受反馈延迟的影响。

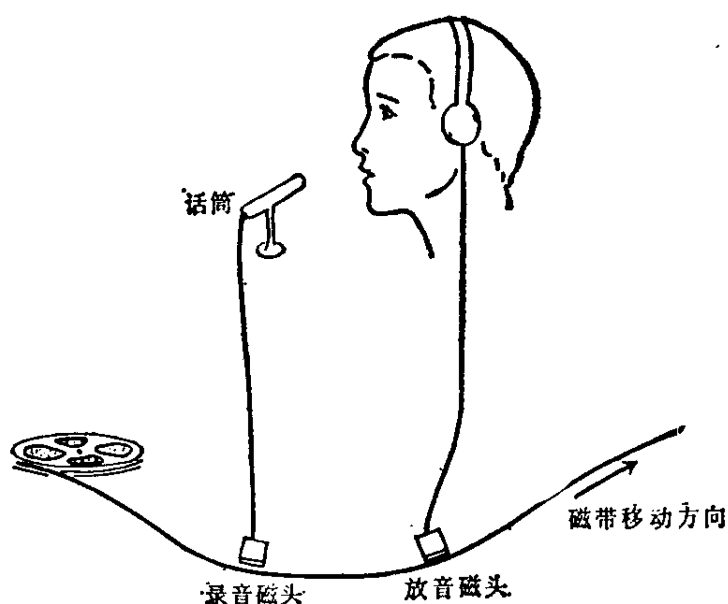


图 2-2

发音器官分析

人类的发音器官是经过几十万年甚至几百万年的演变才成为目前这个样子的。把人类的头部和黑猩猩的头部加以比较，可以看到三点不同。第一，声门的部位不同。黑猩猩的声门部位很高，也就是说，会厌软骨和软腭非常接近(狗的声门部位更高，它的会

厌软骨和软腭不只是接近，而且有很大的交错)。人的声门部位比较低，会厌软骨和软腭相距有几十毫米。这样，人类在发音的时候就可以形成两个气腔，发出的声音种类也就要多得多。黑猩猩由于声门部位高，只有一个气腔，发出的声音种类就少得多。第二，嘴的构造不同。黑猩猩的嘴比较大，因为它要担负很多职能，除了吃东西以外，还要用来打架、移动东西等等。在人类的演化过程中，许多动作都由手来承担了，于是嘴就逐渐缩小，因而也就灵活得多，动作就可以快得多。第三，大脑的大小不同。人和黑猩猩的头部虽然大小差不多，但是人的大脑要比黑猩猩的重两倍到三倍。人类之所以有语言，最大的因素还在于大脑。我们研究人为什么有说话的能力，注意力不能只放在声门以上的发音动作和语音的关系上，还应该多多注意大脑和语言的关系。

用 X 光照相，可以把各种不同元音的舌位清楚地显示出来。[i]、[u]、[a] 这三个元音由于舌面的伸缩或升降，发音器官的形状会发生种种不同的变化。这看起来虽然很复杂，但是可以把它简单化。为了说明发音器官和声波的关系，我们可以把发音器官的两个气腔看成是一根简单的管子。两个气腔在人体里虽然是弯的，但是这对声波来讲丝毫没有关系，所以可以把它看成直的。在不发音的时候，可以看成是一根上下均匀的管子。管子的上端是双唇，下端是声门。发元音的时候，声带是紧闭的，所以管子的上端是开口的，下端是封闭的。发不同元音的时候，两个气腔形状的变化就等于这根管子的不同部分在扩大或缩小。比如，发 [i] 的时候，两个气腔的形状如下页图 2-3 中的上图；如果把它取直，就成为图 2-3 中下图那样一根前细后粗的管子(声门在左，长度以厘米计)。同样，[a] 是一根前粗后细的管子发出来的声音，[u] 是两头粗、中间细的管子发出来的声音。

由此可见，发音器官的形状和不同音色之间的关系并不是不

可捉摸的。发音器官的两个气腔就象是一根管子，管子的形状如果发生改变，发出的声音也必然要跟着改变。发音和声学之间的关系也是很简单的。一根管子发出来的声音，它的频率可以根据管子的形状计算出来。用语图仪分析语音，就等于在测量管子的谐振频率。

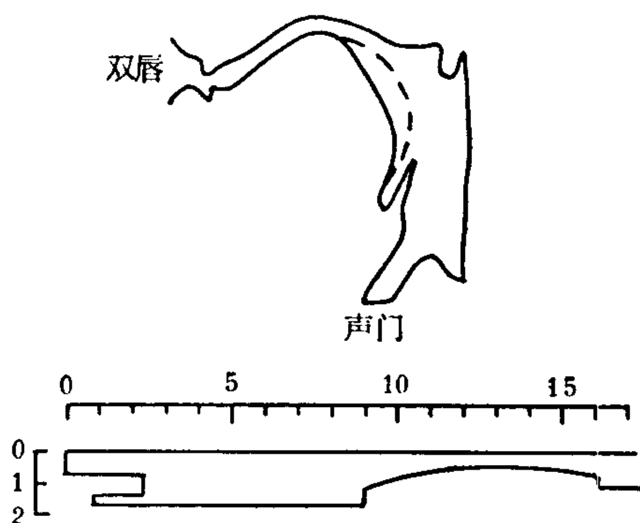


图 2-3

下面专门谈一谈声带。

声带在喉头里边，很难直接观察到它的动作。过去观察声带是用医生用的喉头镜，通过镜面观察声带的动作。这种方法有许多缺点。嘴里放进镜子是很不舒服的。特别是镜子一伸到嘴里去，舌头、口腔就不能活动了，无法发出各种不同的声音。如果光线不够强，还往往看不清楚。最近日本东京医院发明了一种光纤维镜，英文名称是 Fiberoptic, 制造得非常精巧。是很细的管子，可以从鼻子里伸进去，悬在声门的上方。管子能发光，通过它可以观察在正常发音情况下声带是怎样动作的。现在国外语言工作者经常使用它。由于有了这种有效而方便的研究工具，人们现在对声带的动作有了进一步的认识。

声带实际上是非常小的，大约只有十到十三、四毫米长，比小手指头上的指甲盖还要小。呼吸时声带是分开的，发音时声带是合拢的。由于声门下面气流的推动，声带会发生颤动，颤动得快，声音就高，颤动得慢，声音就低。

图 2-4 (见附页)是四对用光纤维镜照下来的声带图，从中可以看出发元音 [a] 时声带的变化情况。每一对图的上面一张，声带

是张开的，下面一张是闭合的。照第一对图的时候，发音人发出来的元音 [a] 是 124 赫，这时声带比较短。第二对图是同一个发音人发的，他把 [a] 发到 174 赫，声带就拉长了，也紧多了。第三对图是 248 赫，频率加了一倍，也就是高了一个音阶。第四对图是发 330 赫的时候照的。从图上可以看到，声音越高，声带拉得越长、越紧，声门也变得越来越细长。如果仔细看的话，还可以从照片上看出声带拉紧以后就变得很薄，放松了就变得相当厚。

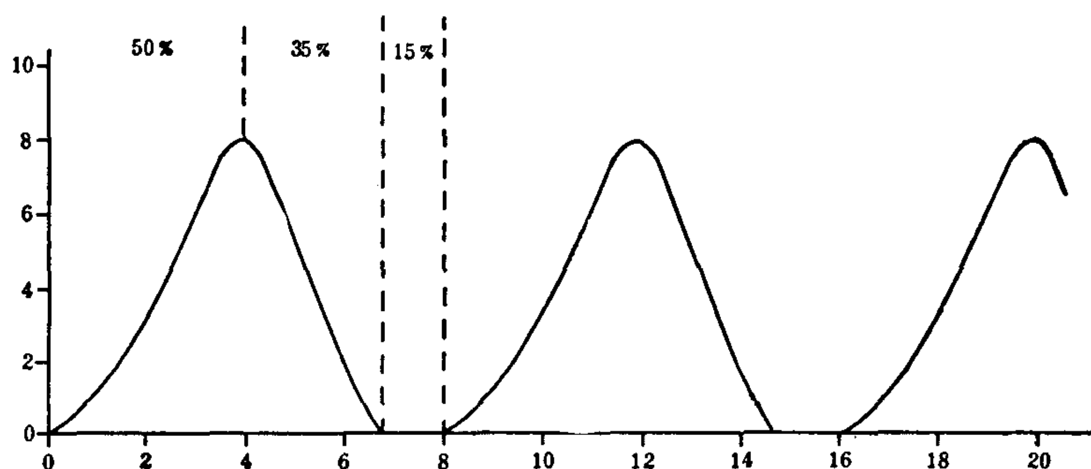


图 2-5

图 2-5 可以把声带开合的动作系统地表示出来。横轴是时间，单位是毫秒；纵轴是左右两个声带当中的面积，也就是它张开的程度。从横轴 0 开始，声带很快地打开。从这张图看，达到最开的时候需要 4 毫秒，这时两个声带之间的面积差不多是 8 平方毫米。到了最高峰以后，声带就开始闭合，速度比张开的时候稍快一点。从开始张开到完全关紧一共有 7 毫秒的时间，关紧以后约有 1 毫秒的停顿，到了第八毫秒又开始打开。声带这种开合变化是周期地进行的。这张图是每隔 8 毫秒开合一次，也就是说，它的周期是 8 毫秒。根据周期和频率的关系（见下一讲），我们可以算出这种振动周期发出的频率是 125 赫。这是一般成年男人的发音频率。

三　语音的物理分析

振幅、周期和频率

语音和其他声音一样，也是由于空气受到干扰发生波动而产生的，所以，除了生理分析以外，还要进行物理分析。这方面的研究完全属于物理声学的范围。

空气的干扰，通常是耳朵感觉得到的。但要是干扰的幅度太低或是频率太高，人的耳朵就感觉不到。感觉不到就听不到声音，从语言的观点看，就是没有语音。一定要有了大脑的反应，才是和语音有关系的声音。

下面先谈谈干扰空气是怎样一种现象。说话人的发音器官和听话人的听觉器官之间充满了无数的空气粒子。发音的时候，比如说发浊音，从肺部出来的气流使声带忽开忽闭地发生颤动。声带张开的时候，空气中的粒子就受到压力向前活动，通常管它叫压紧 (compression)；声带闭合的时候，由于气粒是一种弹性介质，它又向后，也就是朝说话人的方向活动，这种情形，通常叫松开 (rarefaction)。气粒在压紧和松开中左右活动，形成一种疏密波，依次传播，一直影响到听话人耳道里的气粒，引起耳膜的振动，再传达到大脑，就听到了声音。声波的传播是一种纵行的波，气粒的活动方向和波的传播方向是一致的。水波和光波则是一种横波，因为粒子的活动方向和波的传播方向是垂直的。

上面所说的声波活动的特点可以用图 3–1 来表示。图 3–1 里的横轴表示时间，在纵轴上，压紧在上面，松开在下面，用一个气粒代表无数气粒的活动。在空气没有受到干扰的时候，气粒并不动。如果有人说话，这个气粒受到压力的影响，就会活动。压紧的时候

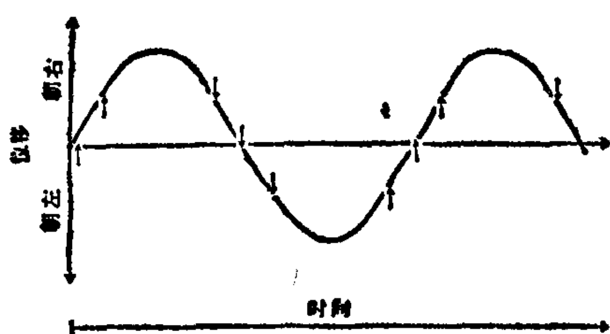


图 3-1

向前活动，画在上面；松开就弹回来，向后活动，画在下面。这一个粒子影响下一个粒子，一个跟着一个地传播出去，一直到能量消耗完了，声波的传播也就停止。

在声波的传播中，我们要注意它的三个物理量。

一、振幅　　说话所产生的气压有强弱的不同，所以压紧和松开的程度也有所不同，它表现为振幅的大小。波幅大，声音就强；波幅小，声音就弱。振幅表示的是音量的大小。

二、周期　　发元音的时候，气粒的活动是跟着声带的颤动按一定时间发生的，它的波形是一种周期波 (periodic wave)。声波在一周期的时间内所传播的距离，叫一个波长。除了元音以外，所有的浊音，如鼻音、浊擦音等等都有周期的部分。清辅音，如 [p]、[s] 等等所产生的波都是非周期的波 (aperiodic wave)。

三、频率　　气粒(或其他物体)在一秒钟内完成全振动(来回一次)的次数，叫作振动的频率。频率的单位是赫兹 (Hz)，也可以写成“周/秒”(cps)。赫兹是德国物理学家 Hertz 的名字，缩写成 Hz，汉语可以简成“赫”。一秒钟振动一个周期就是一个赫兹。

上面谈到了三个常用的基本概念。振幅，通常用 A 来代表，它的单位是分贝(decibel, 经常缩写成db)。周期，通常用 T 来代表，它的单位是秒或毫秒 (ms)。频率，通常用 F 来代表，它的单位是赫兹。

周期和频率有很密切的关系，它们之间是一种倒数关系。这种关系可以有几种不同的代表方法：$FT=1$，$T=1/F$，$F=1/T$。举例来说，如果频率是 50 赫，那么，用 $T=1/F$ 的公式，就可以算出它的周期：$T=1/F=1/50$，$T=0.02$ 秒 $=20$ 毫秒。相反，如果

知道振动周期是 10 毫秒，就可以用公式 $F=1/T$ 算出它的频率：$F=1/T=1/0.01=100$ 赫。

人耳能听到的频率范围是有限的，大致在 20—20,000 赫之间。我们心理上的感觉是频率每增加一倍，音高就增加一个音阶。比方唱歌的时候：$135\dot{1}$，第二个 $\dot{1}$ 的频率是第一个 1 的两倍；第三个 $\dot{1}$ 又是第二个 $\dot{1}$ 的两倍。如果从 20 赫算起：20 赫到 40 赫是一个音阶，40 赫到 80 赫又是一个音阶，80 到 160，160 到 320，每次加一个音阶，这样算下去，最好的耳朵可以听到九个或十个音阶。超出了这个界限，一般人的耳朵就听不见了。上面所说的听觉的频率范围是略去个人的差异说的。通常小孩或年轻人的听觉比较灵敏，可以听到高达 18,000 赫的声音。成年人比较差一些，可以听到 13,000 或 15,000 赫，那要看年龄和健康的情况。老年人往往只能听到 10,000 或 8,000 赫的声音。

跟听觉相比，声带颤动的频率限制就要大得多，一般不会超过 1,000 赫。男人从 60—70 赫开始，最高也许可以达到 200 赫左右。女人声带短而薄，颤动较快，频率大致在 150—300 赫之间。小孩儿更高一些，可以从 200 到 350 赫。这自然是指一般说话的频率，经过训练的歌唱家，唱歌的时候频率可以大大提高。

发音的时候，可以保持声腔的形状不变而只改变声音的频率，这样就形成声调的不同。换句话说，声调就是不同的频率在时间上的变动。下面图 3-2 是我自己发的普通话四个声调的频率

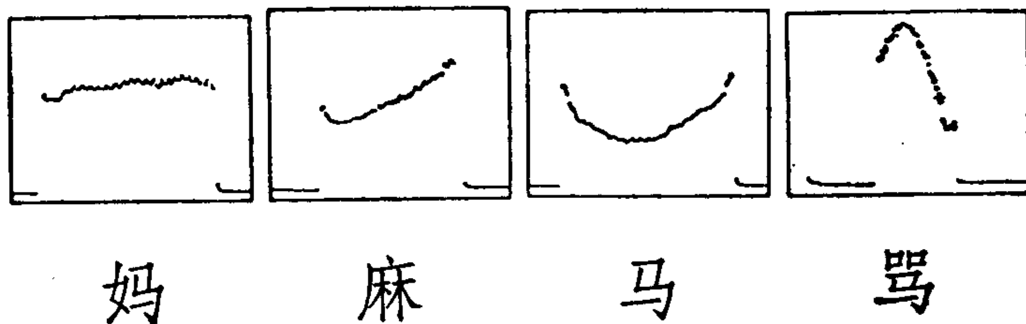


图 3-2

图。声音发出以后，送进电子计算机，马上算出发音的频率，把它画出来，再用照相机把它拍下来，就成为图中的样子。从图上可以看出“骂”一开始相当高，但由于受到鼻音［m］的影响，先产生了一个短短的由低到高的弯头，然后才很快地从最高点 130—150 赫降到大约 90 多赫。由这张图还可以看出，去声是普通话四声中最短的一个。其他三个字“妈”“麻”“马”元音辅音也都是一样的，只是由于声带动作的不同控制造成了发音频率的不同。根据频率测量声调，自然要比现在通行的用五度制定调准确得多，这正是实验语音学对研究汉语方言等等可以做出的贡献。

纯音、复音和谐音

利用音叉，我们可以听到一种非常单调的声音。音叉的振动是最简单、最基本的简谐振动，它发出来的声音只有一种单纯的频率，我们管它叫纯音。世界上的各种声音都是由频率和振幅不同的许多纯音组成的。这种由许多纯音组成的声音叫做复音。在复音中，频率最低、声音最强(即振幅最大)的音叫基音，其他的音都叫陪音(或泛音)，它们的频率都是基音的整数倍，振幅都比较小。通常说某个音频率是多少，都是指基音说的。

一般的声音虽然都是由许多纯音组成的复音，但是我们可以用数学上的傅里叶分析法 (Fourier analysis) 把复音分析成许多纯音。也可以把许多纯音综合在一起成为复音，言语合成就是在这个基础上进行的。傅里叶是一百多年前一位法国科学家，他的贡献之一就是把纯音和复音联系在一起。我们研究语言的人自然不必对他的分析法钻得很深，但是应该从概念上知道这种分析法和语音分析的关系，以及怎样利用它来研究语音。

几何学上的一个最基本的概念是正弦，纯音的声波就是一种正弦波。所谓正弦波，就是用一个角度，把这个角度在单位圈子里

扫过去，(见图 3-3) 经过每个角度时，通过对边和斜边的比例求出它正弦的值来。用公式来表示就是: $f=\sin(\theta)$。角度扫动时是从 0°开始的,从 0° 到 90°,到 180°,到 360°。表现在图 3-3 上,横轴是角度,纵轴是正弦的值(从 1 到 −1)。

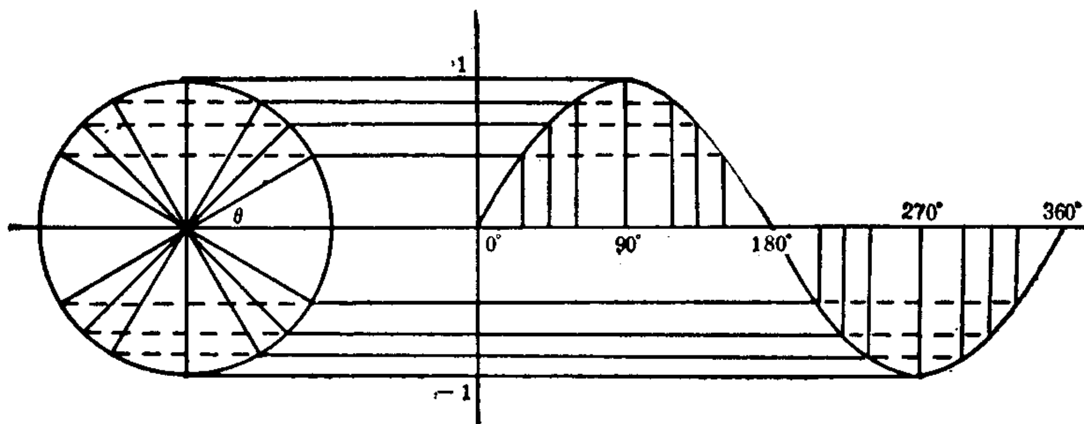


图 3-3

从图上看，θ =0°时，sinθ=0; θ =90°时，sinθ=1; θ =180°时，sinθ=0; θ =270°时，sinθ= −1; θ =360°时，sinθ=0。但这只是大致的情形。如果想知道得比较详细,可以把每个角度的正弦的值都求出来,比如 45°的正弦是 0.70，80° 的正弦是 0.98,等等。这样，把所有的正弦的值都算出来，然后在图上把这些点连起来，所得的曲线就是一个纯音的正弦波形。

一个纯音和另外一个频率相同的纯音是丝毫没有区别的。如果声音里只有纯音的话，那么，两个人用同样的频率和响度说话，声音就应该完全相同;两种乐器演奏同一个曲谱,声音也应该完全相同。可是实际上各种物体发出来的声音千差万别。这是因为这些声音都不是一个简单的纯音，而是包含了许多不同纯音的复音(复合波)。

为了把问题说清楚，我们举一个简单的例子。假定有一个复合波是由一个 100 周/秒的基频加上另外两个纯音组成的，其中一个纯音的频率是基频的两倍(200 周/秒)，另一个是基频的三倍

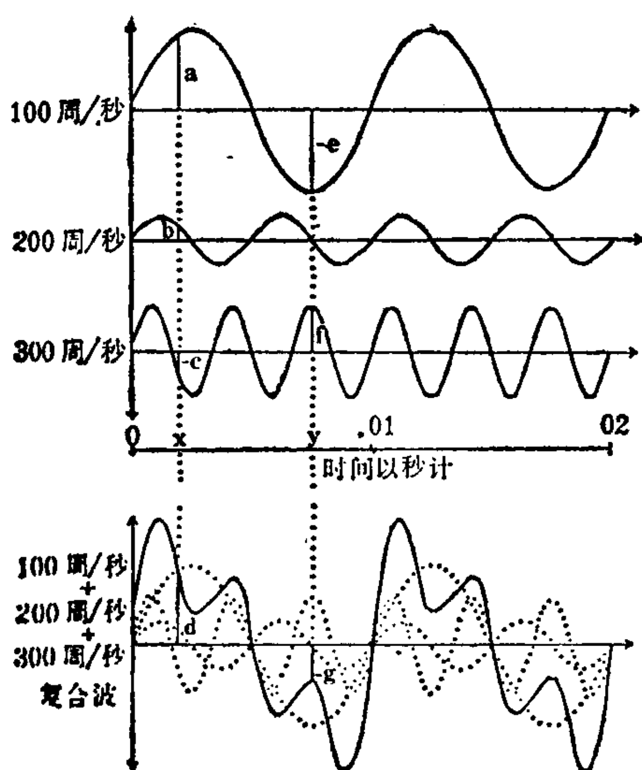


图 3-4

(300 周/秒)，那么就形成图 3-4 那样一个复合波。实际的声音，情况当然要复杂得多。物体所发出的各种声音，除了基音以外，还各自包含着数目不同的陪音，这些陪音的振幅和频率也各自不同，所以才形成了各种不同的声音。

上面所说的这类波形，可以用这样的公式来表示：$f=A\sin\theta$。第二个纯音和它有固定的关系，就是 $f=A_2\sin(2\theta)$。要是还有别的纯音，就可以一直加下去：

$$f=A_1\sin(\theta)+A_2\sin(2\theta)+A_3\sin(3\theta)+\cdots$$

把它缩短，用普通的数学符号表示，就是

$$f=\sum_{i=1}^{n}A_i\sin(i\theta)$$

象这么一串，就叫做“傅里叶系列”。其中最中心的一点，就是在傅里叶系列中，每个单位的频率都有固定的关系，第二个等于第一个的两倍，第三个等于第一个的三倍，第四个等于第一个的四倍。它们的关系可以用一套整数来代表：1，2，3，4，… 可以这样一直加下去。

上面对复音这样分析的结果，还可以用一种叫线谱（line spectrum）的图形来表示。线谱图的纵轴座标还是表示振幅，但是横轴座标表示各个纯音的频率。前面的复音可以用图 3-5 来表示：线谱图上的一条线等于纯音的一个正弦波形，它的高度由振

幅规定(所以也叫振谱)。第一条线代表前面图 3-4 的第一个波形，第二条线代表第二个波形,第三条线代表第三个波形,三条线相加就相当于前面我们加出来的复杂波形。图上第一条线,也就是频率最低的那条,叫做基本谐音，第二条线是第二个谐音,第三条线是第三个谐音,可以有无限之多。要是我们现在分析的是一个周期波,那就不可能在 200和 300 赫之间有一个谐音,因为谐音之间必然是整数的关系， 200 和 300 之间的数和 100 的关系不可能是整数倍。

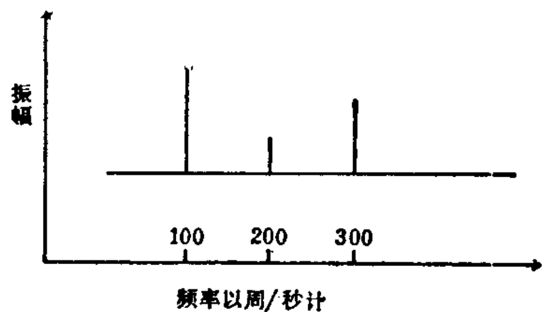


图 3-5

声带的颤动产生声波,声波传到咽腔、口腔和鼻腔后要受到影响。说话的声音就是由这两个不同的功能综合出来的,很难把它们分开。(有位学者用牛做过一个试验: 把牛头砍掉,然后打气,让气流振动声带,这时声带颤动发出的声波就不会受到咽腔、口腔、鼻腔的影响了。)在这种情况下，线谱里很多谐音，它们的振幅总是迅速下降的。谐音的频率越低，振幅越强。由于谐音之间的频率必然是整数关系,因此知道了基本谐音的频率,其他谐音的频率就能确定。例如，要是基本谐音是 100 赫，就可以知道第十个谐音是 1,000 赫;要是基本谐音是 125 赫，第十个谐音就是 1,250 赫。如果突然增加一个音阶，第一个谐音的频率变成 200 赫，第二个就是 400赫,第三个是 600赫,第十个就是 2,000 赫,线谱就稀得多了。如果降低一个音阶，第一个谐音从 100 赫降到 50 赫，第二个就是 100 赫,第三个就是 150 赫，第十个就是 500 赫,线谱就密得多了。从这种变化可以看出一个重要的现象：音高主要由基本谐音决定；知道了基本谐音的频率，其他谐音的频率也就可以确定了。

共振频率和元音音色

频率这个概念在语音上有两种不同的用法，很容易混淆，必须分清楚。由于声带的颤动而产生的频率是主动的频率，用 F_0 代表，F_0 就是声带颤动的频率。由咽腔、口腔、鼻腔的共振产生的频率是被动的频率，用 F_1、F_2、F_3…代表。声带是主动的，因为它自己在颤动。共振的频率和元音、辅音的发音部位有关系，也就是说，它决定于发音器官的形状和体积，所以是被动的。喇叭、笛子、洞箫、黑管都是吹奏乐器，但是音色不同，主要就是因为它们的形状和体积各不相同。[a] 和 [i] 不同，也是由于这个原因。

最简单的管子和瓶子也都有共振。如果想从瓶子和管子的形状或长短算出它的共振频率，还需要知道一个基本概念——波长。在声源和听音人之间的许多气粒中，如果粒子 A 和粒子 B 在时间上的动作完全一样，二者之间又没有别的动作相同的粒子，A 和 B 之间的空间距离就叫波长，一般用希腊字母 λ 代表。波长和频率有固定的关系。如果用 V 代表声音速度，F 代表频率，那么，$\lambda=V/F$。这样，就可以把波长计算出来。假定 $V=340$ m/秒，（声音在空气中传播的速度，正是每秒 340 米）$F=340$ 赫，则

$$\lambda=V/F=340\ \text{m 秒}/340\ \text{Hz}=1\ \text{m}$$

由此可以看出：频率越高，波长越短；频率越低，波长越长。

用同样的方法也可以从管子的长度算出共振的频率，也就是 $F=V/\lambda$。我们曾经把不能主动发音的咽腔和口腔比作一根上下大小一样均匀的管子。一般的成年人，从声门到嘴唇的距离大致是 170 mm。这根管子的长度就是 170 mm。第一个谐音的波长是：$\lambda(F_1)=4L$（L 代表管子的长度）。那么，使用刚才的公式：

$$F=V/\lambda=340\ \text{m 秒}/4\times0.17\ \text{m}=500\ \text{赫}$$

即这根管子的基本谐音的频率为 500 赫。在这种共振情况下的波

形也不是一个单纯的正弦波。它也是复杂的，但是和声带发出来的音的复杂有基本不同。根据傅里叶分析，声带所发出来的声音是用整数来代表的。而上面所说的管子的共振则不同，$F_1=500\ \text{Hz}$，下一个共振 $F_2=3\times500=1,500\ \text{Hz}$，$F_3=5\times500=2,500\ \text{Hz}$，它们之间的关系不是整数 1 2 3 4 5，而是 1 3 5 7 9，是奇数了。我们可以把这套数目简化成下列的公式：共振频率 $Fi=(2i-1)\ V/4L$。这个公式跟 $F=V/\lambda$ 一样，只是稍为复杂一点。根据这个公式，可以推算出：$F_1=1\cdot V/4L$，$F_2=3\cdot V/4L$，$F_3=5\cdot V/4L\cdots$

把咽腔和口腔比作一根均匀的管子是很不自然的，因为发元音时咽腔和口腔的形状很少是那样的。最接近这个均匀的管子所发出的声音是元音 [ə]，[ə] 的 F_1 和 500 赫相近，F_2 和 1,500 赫相近，F_3 和 2,500 赫相近。要比较正确地认识各种元音的不同共振，还需要用各种形状更复杂的管子，象下面图 3-6 这个样子：

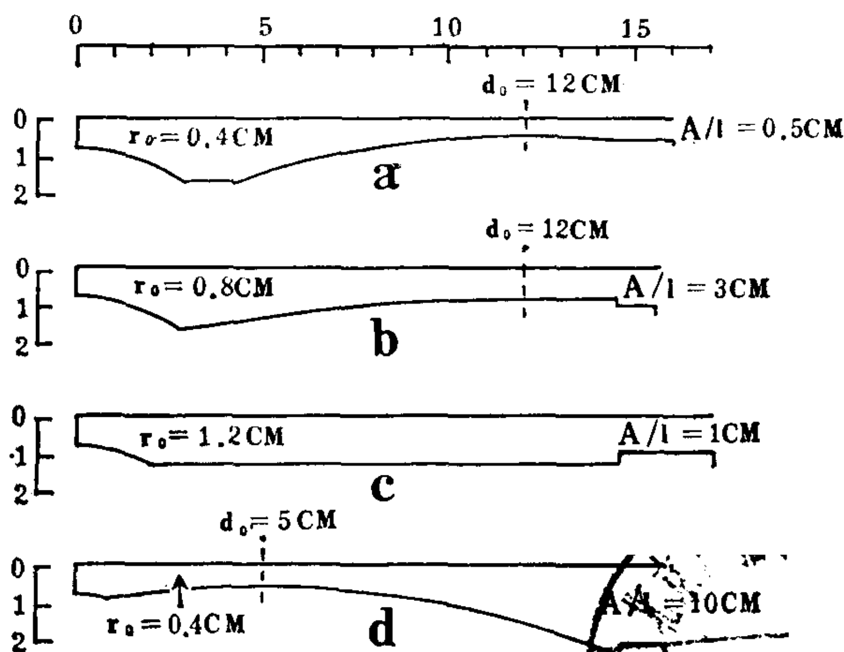


图 3-6

图 3-6 左边从 0 开始是声门的部位。整个管子长 17 cm，代表从声门到嘴唇的长度。管子的各种不同形状代表不同的元音。

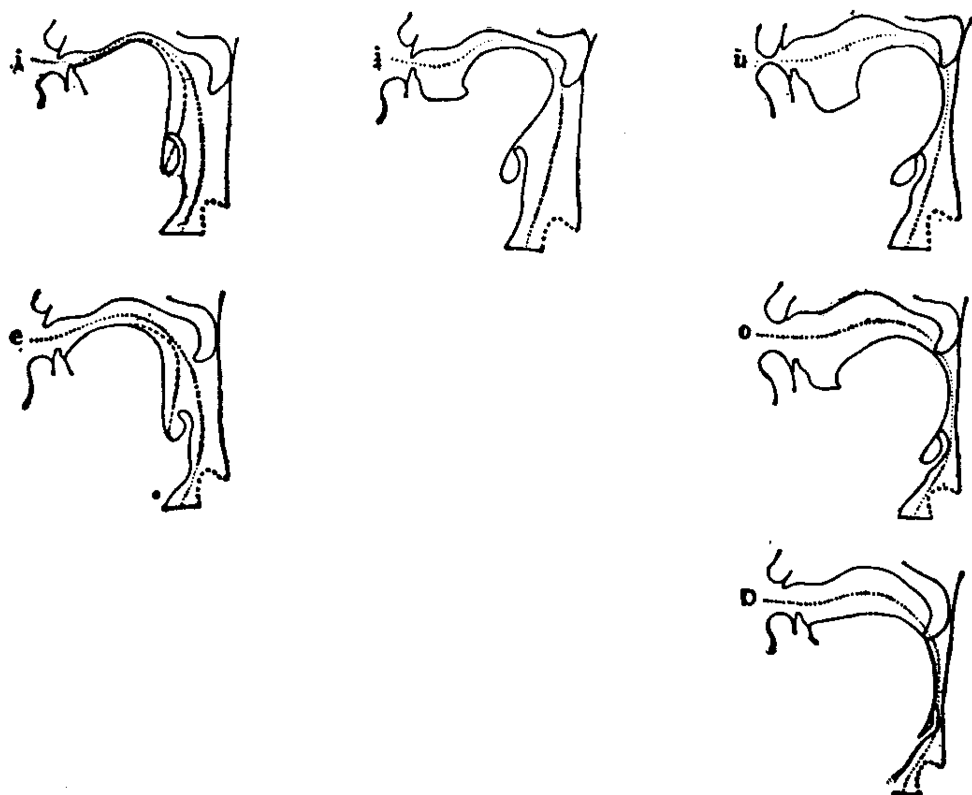


图 3-7

图 3-7 是根据X光照下来的发 [i]、[e]、[ɨ]、[u]、[o]、[ɒ] 时发音器官的形状画下来的。图 3-8（见附页）是发 [ɨ] 时的X光照相，可以参看。如果把这几个元音画成管子，就复杂多了。例如 [i]，管子前部窄，后部宽；[ɒ] 是前部宽，后部窄；[u] 是两头宽，中间窄。从用形状简单的管子到用各种不同形状的管子来代表不同的元音，是最近实验语音学的一个很大的进步，费了将近二十年的时间。

前面已经说到，共振频率、波长和管子的长短这三者之间有固定的关系。我们仍旧用 $F=V/L$ 这个公式来说明。管子长，波长 λ 就增加，λ 的数目大，频率 F 自然就降低；相反，管子短，波长也就短，频率就增高。不同的元音有不同的共振频率，道理也是一样的。比如，发元音 [i] 的时候，嘴唇是向两边展开的，管子就稍微短一些；发元音 [y] 的时候，嘴唇是向前拢圆的，管子就略微长一些。因此，[i] 的共振频率就比 [y] 略高一些。如果连续发 [i-y-i-y]，仔

细去听，我想不需要什么仪器也可以听出 [i] 比 [y] 高一些。

现在我们进一步分析管子内部共振峰之间的关系。先看图3-9 里的四张图。右面这四张图把发元音[i]时声道的复杂形状简单地比作一根管子，右端是封闭着的，左端是开着的。声带颤动，一喷一喷的气流进入咽腔和口腔，就产生了共振，在管子里形成声波的波节(node) 和波腹 (anti-node)。简单一些说，从波节和波腹的关系就可以测量出共振的频率。图 a，表示第一个共振峰，它的频率是 500 赫。图 b，从波形里看到增加了一个波节，位置在整个管长的三分之一的地方，3 × 500赫 = 1,500赫，这是第二个共振峰。下面可以由此类推，用奇数 1,3,5,7……一直乘下去，第三个共振峰的频率是 5 × 500 赫 = 2,500 赫，第四个共振峰的频率是 7 × 500 赫 = 3,500 赫。管子里的波节越多，波长就越短，频率也就越高。

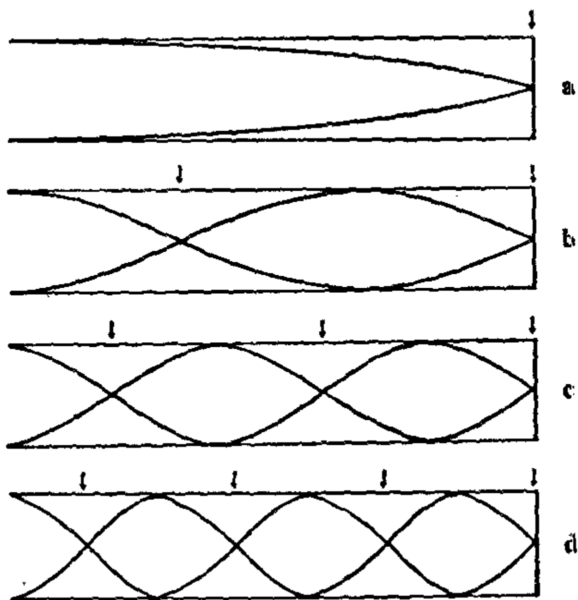


图 3-9

四　语音中的声压和共振峰

语音中的声压

要全面了解共振峰这一基本概念，还需要知道什么是分贝。在传递声音的时候，空气粒子在那里颤动，彼此一会儿压紧，一会儿松开。空气粒子的这种活动，抽象地理解好象很容易，实际去测

量却相当难。我们通常所说的振幅，并不是指单个空气粒子的活动，而是指所有空气粒子合起来的压力。声波的压力叫声压。计量声压的单位叫达因方厘米(dyne per centimeter square，简写为 $dyne/cm^2$)。

达因方厘米是非常小的单位。举例说，如果在手掌上放一块面积两平方厘米的小木片，因为地心吸力的作用，它对手掌可以产生 500 达因方厘米的压力。但是人耳的构造是非常奇妙的，它能感觉到空气中非常非常小的压力，小到仅仅只有 0.0002 达因方厘米的声压；最大则一直可以感觉到 2,000 达因方厘米，不过这时声压太大，耳朵已经非常不舒服，听久了耳膜还会损坏。人所能感觉到的声音，从 0.0002 到 2,000 达因方厘米，其间相差一千万倍。在这样大的范围里，象达因方厘米这样小的单位，使用起来显然是很不方便的。

因此，通常计量声压，不用这个绝对的单位，而用两个数目的比率的对数。用了对数，就可以把很大的数目压缩得相当小。在这方面有一个很常用的概念，叫“信噪比”(signal to noise ratio，简写为S/N，或 SN ratio)。信，指信号，指任何我们想要听到的那个声音；噪，指噪声，指作为信号的背景的那个声音。比如，我想听某人说的话，他的话就是信号；如果这时电扇也在响，电扇的声音就是噪声。但如果我想听的是电扇的声音，电扇声音就是信号，那人说的话就成了噪声。所以信号和噪声都可以是任意指定的某个声音。信噪比就是信号和噪声这两个声音的压力的比率。因为比率的范围太大，可以达到一千万倍，所以用对数比较方便。一般都是用以 10 为底的常用对数，公式是 $S/N=20\log_{10}S/N$。比如 $S=2{,}000$，$N=0.0002$，相差一千万倍，运用公式进行计算，就是：

$$S/N=20\log_{10}2{,}000/0.0002$$
$$=20\log_{10}10^7$$

$$=20\times 7$$
$$=140$$

这个单位就是分贝 (db)。所以，就声压来看，我们听觉的范围是在140 分贝之内，从最微小的声音到最强烈的、会损坏耳膜的声音，范围是 140 分贝。

用分贝来计算声压是很方便的。比如，我把话筒放在离嘴一米远的地方，照平常对话那种响度说话，语音的信噪比大概是 60 分贝左右。如果现在这个房间里的噪声是 1,000 达因方厘米，那就可以推算出我说话的固定数目的声压来:

$$db=20\log_{10} S/N$$
$$60=20\log_{10} x/1{,}000\ \mathrm{dyne/cm^2}$$
$$x=1{,}000{,}000\ \mathrm{dyne/cm^2}$$

信号是噪声的一千倍。如果我用同样响亮的声音打电话，如果电话的噪声比这个房间大十倍，则

$$db=20\log_{10} 1{,}000{,}000/10{,}000$$
$$=20\times 2$$
$$=40$$

信噪比是 40 分贝的电话已经算是不错的了。如果这时要改善信噪比的情况，可以大声说话，增加信号的强度。但是人们在做语音实验时也常常特意把噪声放大，这样可以研究各种语音的不同强度，或人们在不同情况下的注意力，或各别语音的显著性，等等。如果噪声的强度和信号相等，则

$$db=20\log_{10} 1/1$$
$$=0$$

这时信噪比就是零分贝。如果噪声大于信号，比如大两倍，则

$$db=20\log_{10} 1/2$$

差不多是 −6 分贝，变成负数了。近年来有许多语音实验就是在

—6 分贝的情况下做出来的。

我们说话的时候，不同的语音有不同的声压。虽然听起来它们的音强或响度好象相当一致，但如果把说话的声音转化为电流，电压表上的指针就会不断摆动。这就说明：人说话的时候，从一个音到另一个音，声压是在不断变化的。一般说来，语音中浊辅音比清辅音声压大，因为声带颤动时产生较大气压。元音比辅音声压大，因为发元音时气流通道打开，没有很大阻碍，声波的气压较大。由此可以进一步推测出低元音比高元音声压大。比如 [a]，发音时嘴张得很大，口腔前部空间比后部大，咽腔空间很小，整个气腔就象喇叭的样子，所以声压大，可能是所有元音当中声压最大的。[y] 恰恰相反，气腔的前部小，后部大，声压就小。综上所述，我们可以说，[a] 比 [i] 声压大，因为低元音比高元音声压大；[i] 比 [m] 声压大，因为 [i] 是元音，[m] 是辅音；[m] 比 [z] 声压大，因为鼻辅音发音时气流在鼻腔没有受到什么阻碍；[z] 比 [s] 声压大，因为浊辅音比清辅音声压大。总起来就是：[a]>[i]>[m]>[z]>[s]。因此，我们说话的时候，从一个音节到另一个音节，从一个音段到另一个音段，声压总是在变化的。一般用声压级 (sound pressure level) 来说明一个人说话时的声压。但这只是一个平均数，是在他说话的某一段时间内一些声压较高的音相加后的平均数。

不只是音和音之间声压不同，就是一个音内部也有不同声压的部分。一般说来，一个元音可以用三个不同的共振峰来代表，这三个共振峰就有各自不同的声压。

人耳对不同频率的语音的反应能力和声压有密切的关系。如果我们画一个 x-y 坐标图，横轴表示频率，单位为周/秒 (cycles per second，简写为 cps) 或赫兹(赫)，纵轴表示声压，单位为分贝。让振荡器 (osillator) 发出各个不同频率的纯音，在每个频率上逐渐增加它的声压，直到人耳能听见的地步，然后根据它们的赫兹和分贝

画出一条条相应的直线。把所有这些直线的顶端连接起来，就形成一条曲线，叫做等响线 (equal loudness contour)。由图 4-1 中可以看出，人耳对频率低的纯音并不敏感。例如纯音在 40 赫的时候，一定要把声压提高到 60 分贝以上才听得见。频率慢慢增加，人耳也就变得渐渐敏感了。在 2,000、3,000、4,000 赫的区域，最为敏感。这一方面是由人的神经系统的构造所决定的，另一方面也是因为人的外耳本身也是一个管子，它的共振峰频率正好也在这个区域里。超过 4,000—5,000 赫以后，人耳的听觉能力又渐渐降低了。所以图上下面的那条等响线是一条两边高、中间低的曲线。这条线也表示听觉阈 (threshold of hearing)，因为低于这个分贝的音，人耳就听不到了。图上高的那条线表示痛阈 (threshold of pain)，因为这 130 分贝的音，人听久了耳朵就会疼。在这两条线之间，就是人们一般的听觉区域。语音中最主要的一些成分，差不多都集中在这个人耳相当敏感的区域。

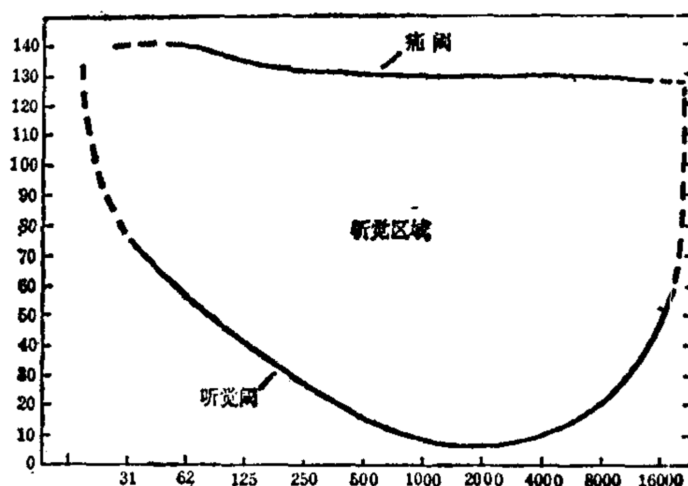


图 4-1

前面介绍线谱的时候，曾经说到由声带发出的声带音的所有谐音，总是频率越高，振幅就越低。所以，在声带音的线谱中，把所有直线的上端连接起来，必然是一条往下降的线，也就是分贝在不断下降。一般说来，在平常发音的时候，声带音的线谱中，频率每升高一个音阶，振幅就差不多降低 12 分贝，也就是 -12db。

带宽和共振峰曲线的计算

首先谈一谈语音的分立和连续现象。

我们在线谱上看到的是一条条垂直的表示谐音的直线。这是因为声带音发音时，声带的颤动是主动的，它推动一个个空气粒子，所以发出的音是一种分立(discrete)现象，就等于是一个一个个别的纯音。可是声带音到了咽腔和口腔，它们对声带音的反应就不是一种分立现象，而是一种连续(continuous)现象了。要表示连续现象，就不能用线谱分析成一条条的直线，而要用一种叫做声谱包络(spectral envelope)的曲线(见图 4-2-1)，它可以象一个套子似的套在线谱上面(见图 4-2-2)。声谱包络说明，在共振峰中心的谐音(如下图中的 250 赫)能量较强，振幅较大，但具有这种谐音能量的并不限于这唯一特定的频率，而是两边都有，只不过离开中心越远，振幅就越低罢了。

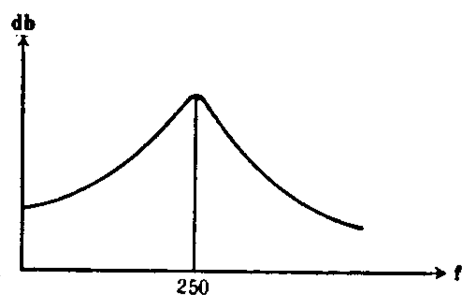


图 4-2-1

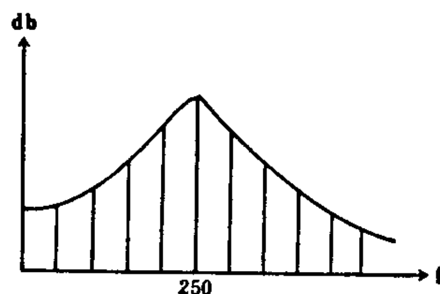


图 4-2-2

分立和连续的这种区别相当重要。因为通常作为研究对象的语音是连续的，而电子计算机所能接受的却只是分立的材料，如果要使用电子计算机来分析语音，就必须使它具有模——数转换(analog to digital, 简称AD)和数——模转换(digital to analog, 简称 DA)的能力，也就是先把输入的语音从连续现象转化为分立现象，进行分析处理，然后再从分立现象转化为连续现象，使分析结果还原成语音输送出来。

声谱包络能表现出共振峰的整个情形。比如一个音的共振峰频率是 250 赫，共振峰的顶端就在横轴 250 赫，左右两边渐渐低下来，左边 249、245、235、··和右边 251、255、270··· 也是有声压的。（见图 4-2-2）如果一个音有好几个不同的共振峰，频率分别是 500、1,500、2,500 赫，（见图 4-3-1,4-3-2,4-3-3）把这三个共振峰合在一起，就成为一个复杂的声谱包络。这个图形就比较接近一般的元音了。（见图4-4）

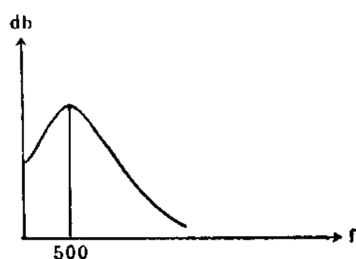


图 4-3-1

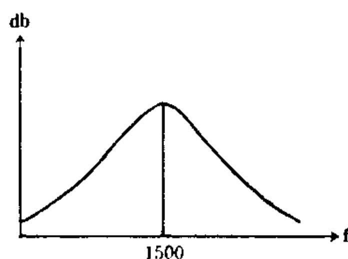


图 4-3-2

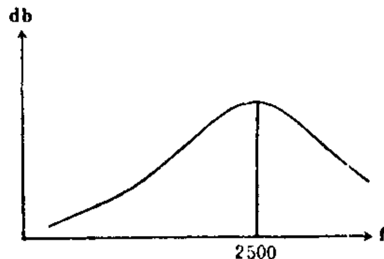


图 4-3-3

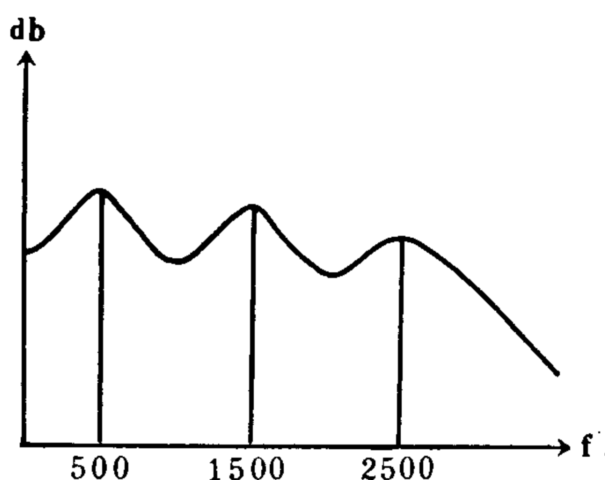


图 4-4

但是，单由频率和振幅这两方面所表示的共振峰，还可以是不同的。在图 4-5中，三个共振峰的分贝相同，频率相同，但波形不同，无法加以区别。这就需要知道每个共振峰的带宽（band width）。带宽这个概念很容易了解：从共振峰的顶点往下降三个分贝（三个分贝的关系就是信噪比$1/\frac{1}{\sqrt{2}}$的关

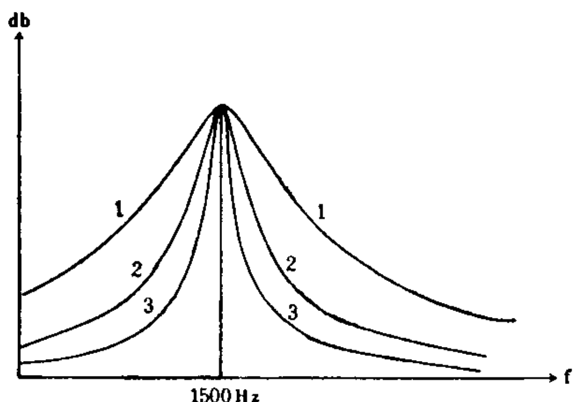


图 4-5

系)，画一条平行线和共振峰曲线相交，相交的两点都有它频率的值，两点间频率之差就是这个共振峰的带宽。比如，左面图 4-6 中共振峰顶点在1,500赫，下降 3 分贝以后，平行线和曲线相交的两点分别在 1,450 赫和1,570赫处，这个共振峰的带宽 B 就是 1,570－1,450＝120 赫。有了带宽，虽然还不能完整地画出不同共振峰的曲线来，但至少可以有一些选择性了，可以大致知道它的形状是什么样的了。

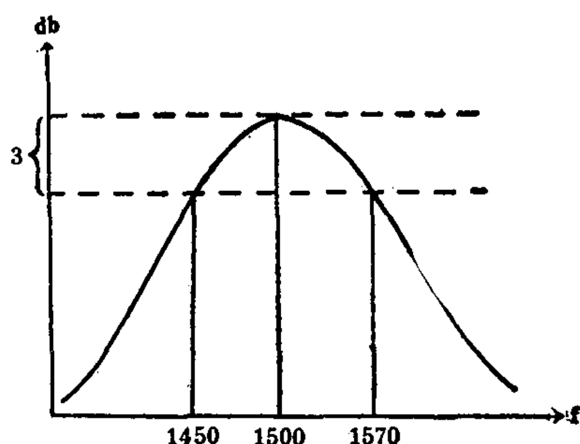


图 4-6

一个音所包含的信息很多，如果用物理方法完整地画出共振峰的曲线，就要对这个音所包含的五六十个谐音中每一个谐音的幅度和相位 (phase) 都加以描述。这样做，既麻烦，又复杂。1956 年，瑞典著名语音学家 G. Fant 写了一篇重要的文章，叫《从频率预测共振峰声级和声谱包络的可能性》(On the Predictability of Formant Levels and Spectral Envelopes from Frequencies)。他在这篇文章中提出了一种可以仅仅根据共振峰的频率和带宽来计算共振峰声谱包络的简便方法。通常共振峰的频率范围很大，可以低到 200 赫左右，高到好几千。而带宽的范围则很窄，最低差不多 40 赫，最高差不多 250 赫，平均在 100 赫左右。现在假设一个共振峰的频率是 250 赫，带宽是 100 赫，按照 Fant 的方法画一个复数平面图，(见下页图 4-7) 横轴表示频率，纵轴表示带宽。纵轴只画带宽 (B) 的一半：$B/2$，在这里也就是 50 赫(100/2)。从 $B/2$ 的地方画一条平行线和横轴上 250、－250 两处的两条垂直线相交，构成一个长方形。然后画几个矢量 (vector)，—— Fant 在这里用

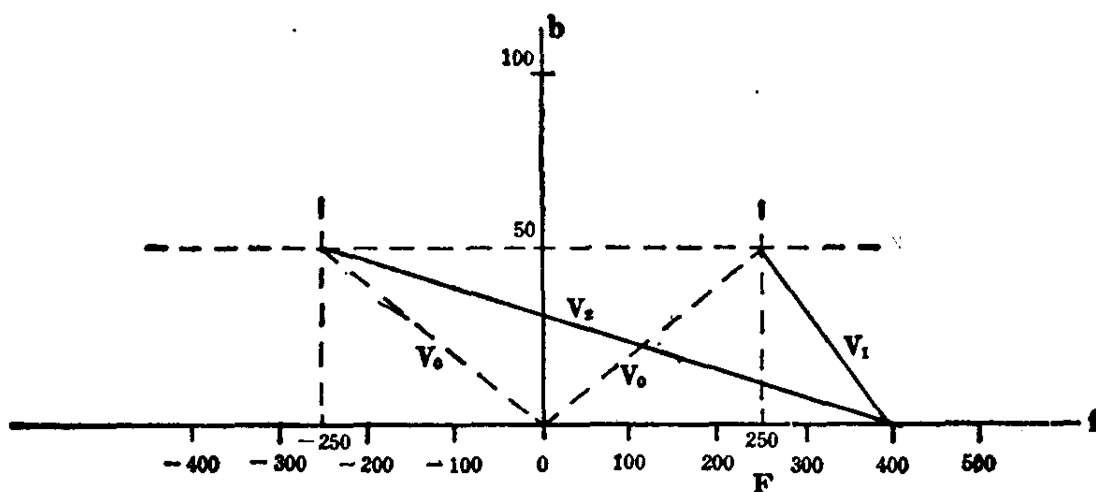


图 4-7

了 vector 这个词，应该译成“矢量”，但是实际上它比真正的矢量简单一些，因为它只有长度，没有角度，——从这几个矢量的值就可以求出各个共振峰的曲线。

下面简单介绍一下计算方法。

首先由零点向 250 和 50、−250 和 50 这两个相交点画出两个相等的矢量 V_0。如果想要知道共振峰在比如 400 赫的时候有什么值，就再从 400 赫的 f 点向这两个点画出两个不等的矢量 V_1 和 V_2，求出来的分贝值就是：

$$H(F)=20\log_{10}\frac{V_0V_0}{V_1V_2}$$

V_0、V_1、V_2 都是直角三角形的斜边，斜边 $C=\sqrt{a^2+b^2}$，按照公式推导：

$$V_0=\sqrt{F^2+\left(\frac{B}{2}\right)^2}$$

$$V_1=\sqrt{(f-F)^2+\left(\frac{B}{2}\right)^2}$$

$$V_2=\sqrt{(f+F)^2+\left(\frac{B}{2}\right)^2}$$

则

$$H(f)=20\log_{10}\frac{F^2+\left(\frac{B}{2}\right)^2}{\sqrt{(f-F)^2+\left(\frac{B}{2}\right)^2}\cdot\sqrt{(f+F)^2+\left(\frac{B}{2}\right)^2}}$$

求出来的就是 f 点上分贝的单位。

这样可以把横轴上每一点的 $H(f)$ 都计算出来，也就是说，可以把整个声谱包络的形状画出来。不过这样太复杂了。根据 Fant 的方法，可以把它简化，以便找出其中各种内部关系。比如，在几个数值中，如果相对说来 F 值相当大，B 值很小，就可以把 B 略去不计。然后再看下面几种情况:

1. 如果 $f=F$，则

$$H(f)=20\log_{10}\frac{F^2}{\frac{B}{2}\cdot 2F}$$

$$=\frac{F}{B}$$

这时共振峰频率越高，包络的振幅也越大。比如低元音 [a] F_1 的频率（约 700 赫）比高元音 [i] F_1 的频率（约 250—270 赫）高，振幅也比 [i] 大。

我们可以在上述方程式中代入数字。设 $F=250$，$B=100$，在 $f=F$ 的时候，则

$$H(f)=20\log_{10}\frac{62500+2500}{\sqrt{2500}\cdot\sqrt{250000+2500}}$$

$$=20\times 0.41$$

$$=8.2$$

这就是共振峰峰顶的分贝数。减去了 3 分贝后，带宽所在的两个交点应该是 8.2－3＝5.2db。

2. 如果 $f>F$，例如 $f=300$，则

$$H(f)=20\log_{10}1.66$$

$$=4.4$$

这个值比 5.5db 要小，从而使曲线两边不对称。（见图 4-8）这是因为人的咽腔、口腔并不是完全均匀的，所以发出的声音的共振峰也不对称，往往是一面宽，一面窄。宽的那面声压降得慢，200 赫时是 6.1 分贝；窄的那面降得比较快，300 赫时是 4.4 分贝。

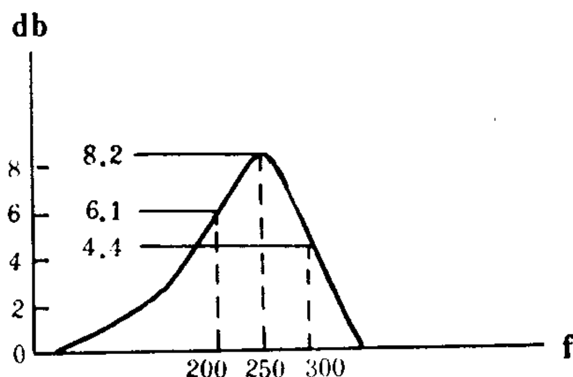


图 4-8

3. 如果 f 相当大，则可以再简化成为

$$H(f)=20\log_{10}\frac{F^2}{f^2}$$

这时 f 每增加一倍，振幅的变化就是

$$H(f)=20\log_{10}\frac{1}{4}$$
$$=-12\text{ db}$$

也就是振幅降低 12 分贝。元音的共振峰中，F_2 F_3 象是骑在 F_1 背上似的。F_2 F_3 离 F_1 越近，振幅就越高；离 F_1 越远，振幅就越低。离 F_1 的距离每增加一倍，振幅就降低 12 分贝。

下页图 4-9 是电子计算机根据 Fant 的公式计算出来的共振峰。小圆圈是不同的 f 的值。把这些小圆圈连接起来，就成为一个完整的共振峰的图形。

计算共振峰的声谱包络，还要考虑两个因素。一是 F_4 F_5 等等更高的共振峰。（一个音的共振峰可以有几十个，能够测出多少，要看仪器的性能。）而且即使我们不去计算 F_4 F_5 等等，也至少要考虑到它们的影响，把它们看作一个因素，也就是在频率越高的地方越要考虑增加振幅。二是辐射的分布。话筒接受到的音要受到话

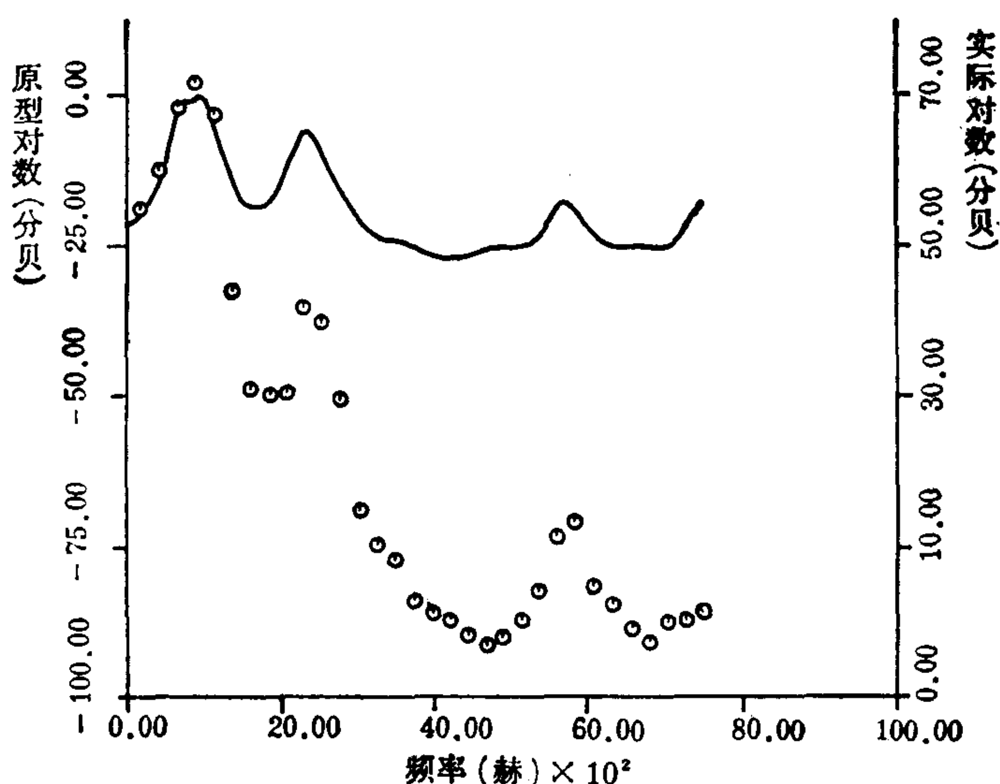


图 4-9

筒和发音体距离的影响。如果距离越来越近，频率的包络就会改变，通常是每提高一个音阶增加 6 分贝。图 4-9 中，上下两个包络的共振峰频率一样，可是振幅不一样，原因就在于上面的一条是把这两个成分算进去了，而下边的一条没有这样考虑，因此频率高的时候振幅就低得多了。

五　语音的基频、共振峰和元音的关系

基频和共振峰频率的关系

发元音的时候，声带颤动产生频率 F_0，发音器官的谐振形成共振峰 F_1 F_2 F_3 等等。F_0 是完全由肌肉、软骨的各种动作支配着发出来的，是主动的；F_1 F_2 F_3 等等的频率则是由发音器官的形状所决定的，是被动的。这些被动的共振峰之间相互有影响，但它们

和 F_0 则是相互独立的，并没有相互依存的关系。比如人们说话的时候可以不让声带颤动，发成耳语音，这时并没有 F_0，但还是可以有 F_1 F_2 F_3 等等。所以，声带颤动时是 [i] [a] [u]，声带不颤动时仍然可以听出是 [i] [a] [u]。这是由于发音器官的形状不变，共振峰的频率也就不变，即使声带不颤动，仍然可以分辨出这些信息。

如果发音器官的形状不变而 F_0 频率改变，语音的音高就会发生变化。比如汉语普通话的“衣” [i⁵⁵]，“移” [i³⁵]，“以” [i²¹⁴]，“意” [i⁵¹]，念这四个字时共振峰的频率没有改变，但是 F_0 变了。音高在任何一种语言里都是有作用的。世界上每种语言都有语调，语调的高低就是由 F_0 的频率所决定的。但是在一些语言里，比如汉语，音高还有双重的作用，也就是说，不仅有语调，还有声调（字调）。

声调在耳语的时候也能表达，这道理似乎有点讲不通，因为前面提到声调是由 F_0 传达的，而耳语时恰好声带不颤动，没有 F_0。但是，一般语音中某个成分的表达事实上都要依靠几个方面，并不单独依靠一个方面。声调也是如此。它不只靠 F_0，还靠振幅，长短，等等。以汉语普通话上声字“马”为例。如果以时间为横轴，F_0 为纵轴画一个图，（见下图 5-1-1）又以时间为横轴，分贝为纵轴画一个图，（见下图 5-1-2）两个图形是差不多的。

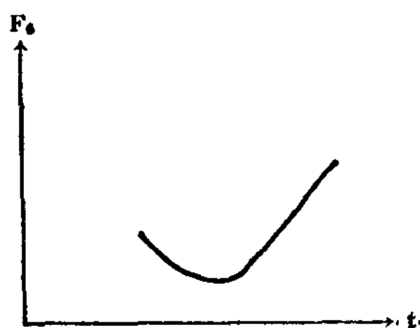


图 5-1-1

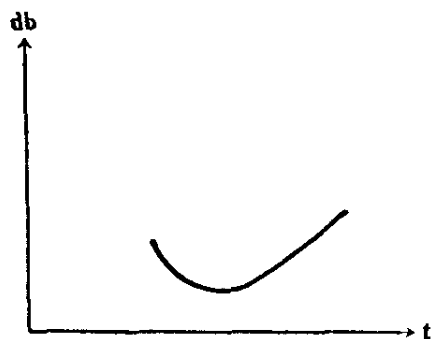


图 5-1-2

十几年前我在俄亥俄(Ohio) 州的时候指导过一篇博士论文，讲的是汉语普通话音节的辨别，其中谈到在声调里 F_0 和振幅的关系。通过实验发现，区别四声，单靠 F_0，成功的可能性在百分之八十以上；单靠振幅，成功的可能性也在百分之七十以上。由此可见，耳语中声调的表达，实际上是有振幅在起作用的。

如果不改变 F_0 的频率，而改变共振峰的频率，也就是改变发音器官的形状，语音的音色就会发生变化。如果在同一音高上先后发出元音[i-y-a-ə-o]，就有可能听出其中起改变语音音色作用的单个的共振峰。我们在听人说话的时候，通常只是为了了解语义，只把语音当作语义的代表来看待，对语音本身并不很注意，因此很难听到它中间的不同部分。但如果不考虑它所担负的语义，把注意力集中在语音本身，就有可能听到这些部分。比如我们知道，前元音的 F_2 的频率总是比后元音的高得多(见下文)，如果我们发一个前元音 [y]，然后在发音过程中舌位逐渐后移变成发后元音 [u]，同时音高一点不改变，这时就能听到一个尖锐的、类似口哨那样的声音在降低，这就是 F_2 在降低频率。如果再从 [u] 变到 [y]，就可以听到这个声音又在升高，这就是 F_2 的频率在升高。因为 F_2 的频率范围变动最大，所以比较容易听到它。

元音基频和共振峰频率、振幅的值

在五十年代，G. E. Peterson 和 H.L.Barney 曾经做过美国英语元音中基本材料的测量工作。他们根据七十六个不同性别、年龄的发音人提供的语音材料，分析综合，进行实验。从四十年代开始研究，到 1952 年发表了一篇文章，叫《控制方法在元音研究中的应用》(Control Methods Used in a Study of the Vowels)。他们在这篇文章里列出了下面这个表格：

	i	ɪ	ɛ	æ	ɑ	ɔ	ʊ	u	ʌ	ɚ
基频(Hz)										
男	136	135	130	127	124	129	137	141	130	133
女	235	232	223	210	212	216	232	231	221	218
儿童	272	269	260	251	256	263	276	274	261	261
共振峰频率(Hz)										
F_1 男	270	390	530	660	730	570	440	300	640	490
女	310	430	610	860	850	590	470	370	760	500
儿童	370	530	690	1,010	1,030	680	560	430	850	560
F_2 男	2,290	1,990	1,840	1,720	1,090	840	1,020	870	1,190	1,350
女	2,790	2,480	2,330	2,050	1,220	920	1,160	950	1,400	1,640
儿童	3,200	2,730	2,610	2,320	1,370	1,060	1,410	1,170	1,590	1,820
F_3 男	3,010	2,550	2,480	2,410	2,440	2,410	2,240	2,240	2,390	1,690
女	3,310	3,070	2,990	2,850	2,810	2,710	2,680	2,670	2,780	1,960
儿童	3,730	3,600	3,570	3,320	3,170	3,180	3,310	3,260	3,360	2,160
共振峰幅值(db)										
L_1	−4	−3	−2	−1	−1	0	−1	−3	−1	−5
L_2	−24	−23	−17	−12	−5	−7	−12	−19	−10	−15
L_3	−28	−27	−24	−22	−28	−34	−34	−43	−27	−20

这个表格也许是实验语音学中最有名的一个表格了。从表上可以看出，女人发音的频率比男人高，儿童更高。比如表中元音 [i] 的基频，也就是 F_0，男人是 136 赫，女人是 235 赫，高出 100 赫左右，十岁的儿童是 272 赫，更高。元音中共振峰的频率，越往上越高。比如元音 [i]，男人发音，F_1 是 270 赫，F_2 是 2, 290 赫，F_3 是 3, 010 赫。共振峰的振幅在表中用 L 代表，L_1 就是第一共振峰 F_1 的分贝值，L_2 是 F_2 的，L_3 是F_3 的，各有它的特征。总的来说，频率越低的共振峰，振幅就越大。所以 L_1 总是比 L_2 大，L_2 总是比 L_3 大，低元音共振峰的振幅也总是比高元音的大。在所有单个的共振峰中，元音 [ɔ] 的第一共振峰 F_1，由于有一个最靠近的 F_2 的影响，振幅最大，因此被当作标准，把它定为 0，所有其他共振峰的振幅都和它比较着来定值。每一个共振峰的振幅并不只是由它自己的频率决定的，共振峰和共振峰的距离越近，振幅就越高，距离越远，振幅就越低。例如 [æ] 和 [ɑ] 的 F_1 的频率虽然比 [ɔ] 高，但离开 F_2 比较远，所以振幅反而不如 [ɔ] 的 F_1 高。振幅最低的是 [u] 的 F_3，降低到了 −43 分贝，这一则是因为它的 F_1 比较低，(高元音的 F_1 总是比较低的) 二则是因为它的 F_2 和 F_3 之间的距离大。

共振峰频率 F_1 F_2 和元音舌位的关系

有人把 Peterson 和 Barney 的表格加以简化，画成另一种图。(见下页图 5-2) 横轴表示八个元音，纵轴表示共振峰的频率。从图上可以看出，F_1 的位置从 [i] 到 [ɑ] 逐渐上升，从 [ɑ] 到 [u] 又逐渐下降；F_2 则从 [i] 到 [ɔ] 一直急速下降，从 [ɔ] 到 [ʊ] 或 [u] 又略有上升；F_3 从 [i] 到 [u] 虽然也是一直下降的，但下降得非常缓慢。

如果以共振峰的 F_2 为横轴，F_1 为纵轴，零点放在右上角，就

可以画出一个元音图，(见图 5-3) 通常叫做共振峰图解 (formant plot)。它和传统语音学中的元音舌位图很相似。把它和元音舌位图比较，可以看出两点：首先，元音舌位的高低跟 F_1 有关，舌位越高，F_1 就越低，舌位越低，F_1 就越高，二者存在一种相反的关系。其次，元音舌位的前后跟 F_2 有关，由 [i] 到 [u]，F_2 逐渐降低，反过来由 [u] 到 [i]，F_2 逐渐升高。一般讲来，在共振峰图形四周

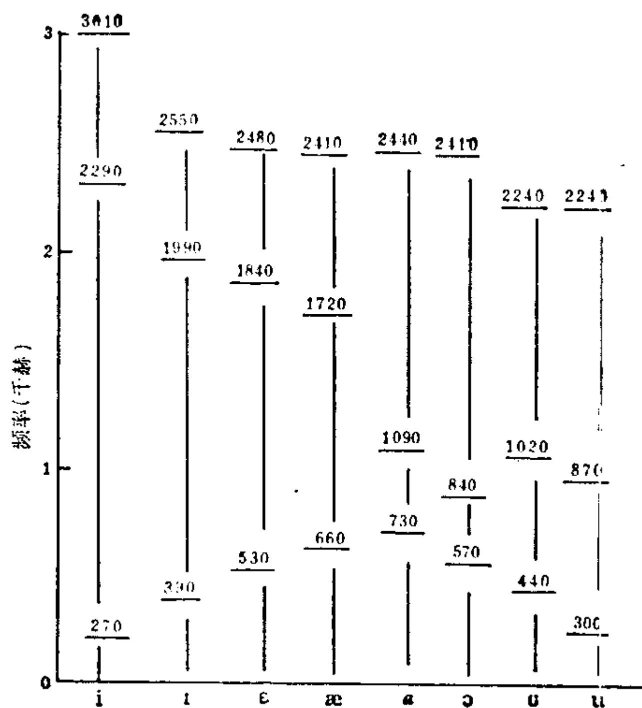


图 5-2

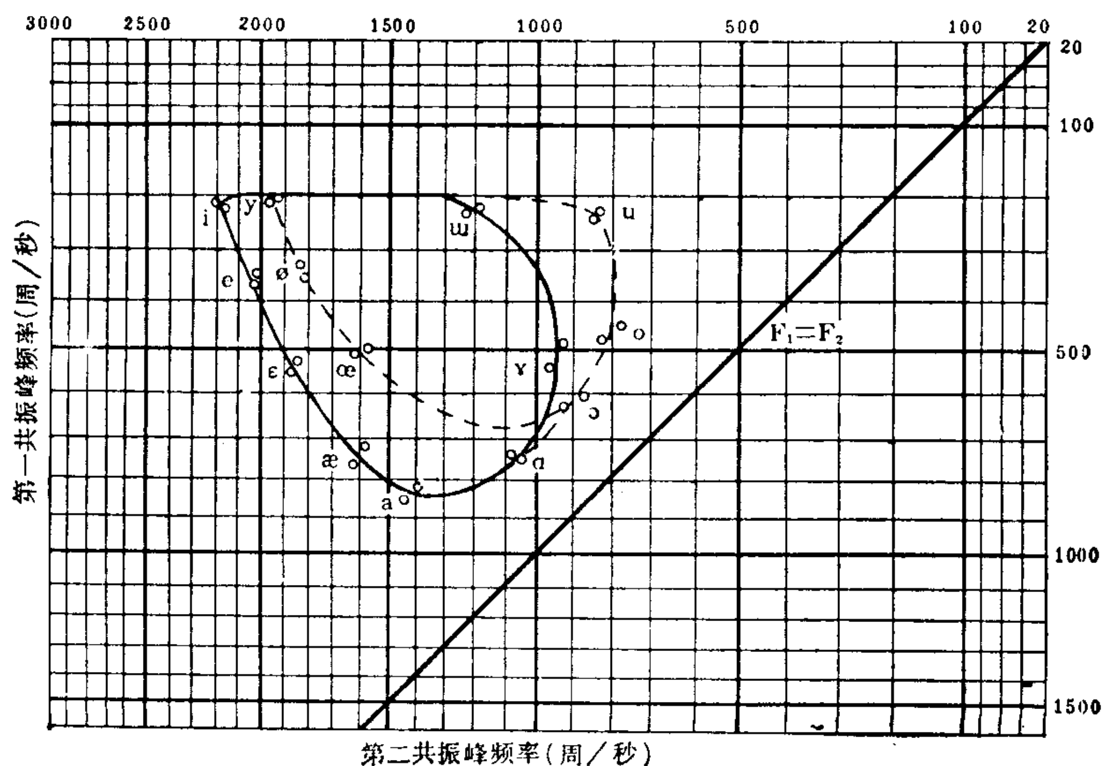


图 5-3

的音如 [i] [u] 等，听起来比较清楚，在圈子中间的音听起来就比较难辨别。只有一个 [ɚ] 是例外，如果画出来，它虽然也在中间，但也可以听得很清楚，这是因为它的 F_3 降得很低，有它特别的地方。

共振峰图解所用的刻度 (scale) 目前还不是统一的，大约有四种。第一种是用整数作底的刻度。第二种是用对数作底的刻度，通常是用以 10 为底的对数，即常用对数。第三种叫 Koenig 刻度，是一个叫 Koenig 的工程师发明的，这种刻度是从 0 到 1,000 赫用整数，1,000 赫以上用对数。第四种比较难用，但是最合理，叫嗅刻度 (mel scale)。这是根据人耳对纯音高低的反应而设计的。比如现在画一个图（见图 5-4），横轴是频率，从 0 标到 10,000，纵轴是嗅，从 0 标到 3,000。从图中可以看出人耳对音高的反应：1,000 嗅是 1,000 赫，2,000 嗅差不多是 3,500 赫，500 嗅大致是 400 赫左右。可见嗅和频率的关系并不很简单，这是因为人耳的构造不完全是均匀的。

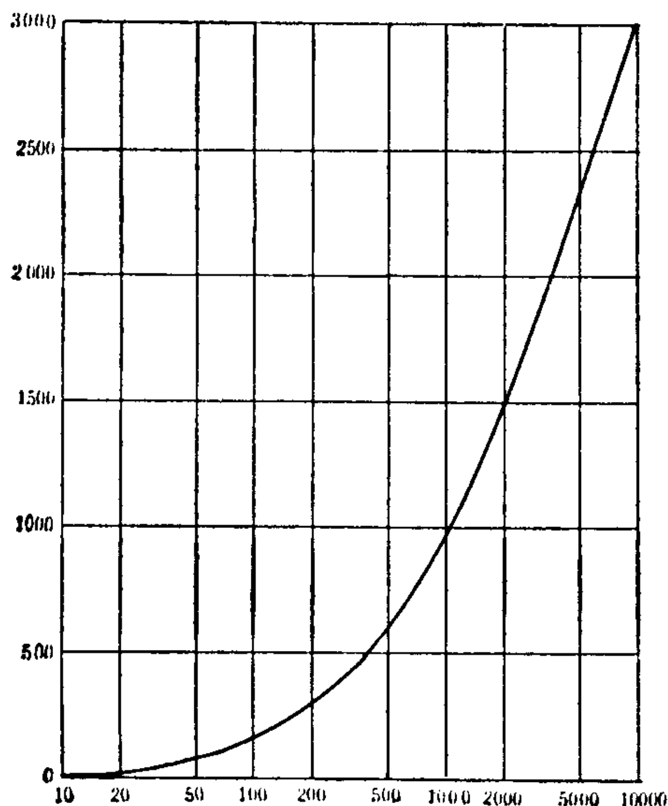


图 5-4

从以上所谈可以看出，元音既可以根据舌位的高低前后来分类，也可以根据共振峰频率的位置来分类。共振峰图解虽然比较后起，但实践证明它似乎比较可靠，因为共振峰的频率是可以客观地测量出来的。元音舌位图虽然是我们早就熟悉的，但

舌位的高低前后其实倒是一个相当模糊的概念。比如根据X光照相画出来的元音舌位图，元音 [i] 仔细看来并不见得就一定可以算是一个前高元音。又何况不同的人发音，舌位会有很大差别，就是同一个人在很短时间里发出几个相同的音，舌位也不会完全相同。传统语音学认为元音分类的基础是生理上的，根据目前的了解，这个基础似乎不是生理上的，而是声学上的。

但是，按照声学特征来对元音进行分类，也还是一个很复杂的问题。请看下面图 5-5：

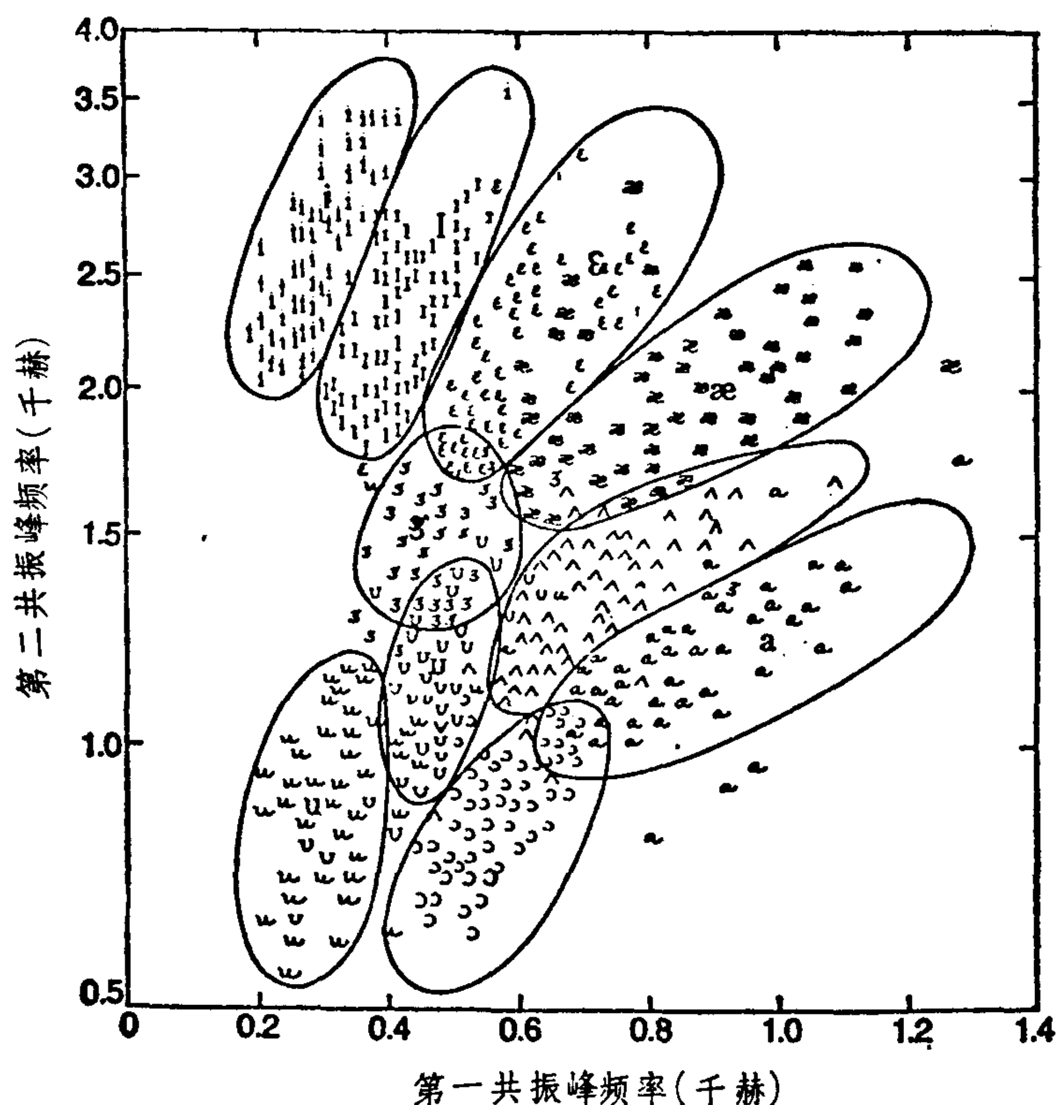


图 5-5

这个图也是 Peterson 和 Barney 根据实验结果画出来的。

图中以 F_1 的频率为横轴，用线性刻度；以 F_2 的频率为纵轴，用 Koenig 刻度。中间画的是七十六个发音人所发的每一个元音的 F_1 和 F_2 的值。结果每一个元音都形成一个椭圆形的分布范围。问题复杂的是在一个椭圆形里有时并不只有一种元音，比如在 [ε] 椭圆形里还有 [ɪ] 和 [æ]。这就是说，如果仅仅考虑 F_1 和 F_2，同一个元音听起来还可能有时是 [ɪ]，有时是 [ε]，有时又是 [æ]。而事情显然不应当是这个样子的。所以还必须考虑到 F_1 F_2 以外的成分。事实上，这种复杂现象之所以会产生，正是因为语音中的成分不止有 F_1 和 F_2。除 F_1 F_2 以外，还有 F_3 和 F_4 等，都是有它的重要性的。仅仅提出 F_1 F_2，不过是一种简单的代表方式。所以近年来许多语音学家尝试用种种不同的图解方法，有时是用 F_1 和 F_2，有时是用 F_1/F_3 和 F_2/F_3，有时把 F_0 也加进去，这样，图也就越画越复杂了。显然，考虑到的因素越多，语音分析获得成功的可能性也就越大。如果只用 F_1 F_2，那是根本不够的。

对听到的语音进行判断时，心理方面所起的作用也许更为复杂。P. Ladefoged 和 D. E. Broadbent 曾经在这方面做过一个实验，写在《元音传递的信息》(Information Conveyed by Vowels)一文中。他们试图解答这样的问题：一个人听到一个元音，怎么就能知道那是某个元音？单凭 F_1 F_2，或再加上 F_3，是不是就能让他作出判断？当然，这可以从物理方面来考虑，——知道了一个元音的 F_0 和 F_1 F_2 F_3 等，耳朵就可能告诉大脑那是个什么元音。但是他们有另一种想法，认为决定一个元音的不仅仅是物理上的条件，还应该看那个音附近的环境，因为人们的听觉是受周围环境影响的。事实上，一个男人和一个女人都发一个元音 [æ]，共振峰频率自然差得很远，但听起来仍然都是[æ]。

他们首先选择下面四个元音不同的词：

bit　　bet　　bat　　but

[ɪ]　　[ɛ]　　[æ]　　[ʌ]

然后用语音合成器合成这四个词，由机器发出音来让几十个人听。在这种情况下，每个人都能把这些词的元音区别得很清楚，判断也相当一致。然后他们又选择"please say what this word is"（请说出这个词是什么）这样一个句子。这个句子也有不同的元音，如[i]、[eɪ]、[ɐ]、[ɚ]、[ɪ]等。他们有系统地分六种情况改变这些元音的共振峰频率。同是这句话，让机器发出元音音色有不同变化的六句话来。把这六句话和 bit 等四个词任意配合，让机器先念句子后念词，再让听音的人分辨这四个词。这时情况就变得完全不同了。同是一个 bit，在这个句子后面听起来是 bit，在另一句子后面听起来就成了 bet。同样，原来的 bet 有时听成了 bat。这样，词里的元音就变得要根据前面句子里元音的值来决定了。他们实验的结论是，人在听话的时候，要判断听到的是哪一个音，这并不是绝对的，而是相对的，也就是说，并不是完全根据物理条件来决定的。

从三十年代开始，很多人曾经希望能使用一种能听懂语音的机器，并且进行了大量实验。这种研究工作叫"言语识别"（speech recognition）。美国许多大公司也希望能最终制造出一种能听懂人说话的机器，并且为此大量投资。但是几十年来，花费了几十万美元，收获却不大。当然，我们今天对语音的分析是进步多了，声谱图给我们提供了很多新的知识，但是要让电子计算机完全能听懂人说话，至少在目前还仍然等于是一种梦想。电子计算机听音确实非常准确，至少不比人耳差，但是它仍然不能象人一样完全听懂别人说话。这是因为人听话判断语音的时候，不是单靠其中的物理条件，而是还要加进许多有关的成分，比如，用脑子这个成分，而到现在为止，我们还一点不了解人的大脑为什么会具有这种能力。

六　语音的外来变化和内在变化

两种不同的语音变化

在人们自然说话的时候，连在一起的两个音会因为相互影响而发生变化，有时是同化，有时是异化。这种变化，由于它是由语音的外部环境造成的，我把它叫做外来变化 (extrinsic variation)。这种变化是传统语音学几百年来一直在注意的，例子也很多。比如，[lin] 里的 [i] 和 [li] 里的 [i] 相比，总有点不同，前一个 [i] 略微有一点鼻化。这是因为它后面有一个鼻音，发音时软腭提前下垂，就使这个元音鼻化了。元音鼻化，象现代法语里那样，是几乎每一种语言历史上都有过的现象。再举一个例子。英语 lull [lʌɫ] 里有两个边音，前一个是舌头前部靠近硬腭，后一个是舌头后部靠近软腭，这两个边音的发音部位是不能倒换的。还有一个常见的外来变化的例子。在比如英语、德语、俄语这样一些语言里，音节开头的清塞音一般是要送气的，如 par [p‘a:]，但如果前头有个清擦音，就变成不送气了，如 spar [spa:]。外来变化一般只和特定的语言、特定的时代有关系。现代英语有这种现象，二百年后也许就没有了。有的语言则根本就没有过这种现象。

和外来变化相对，还有一种语音变化，我把它叫做内在变化 (intrinsic variation)。内在变化可以从两个方面来讨论。首先是把它看成一种物理现象，用声学仪器按声学的公式对它进行种种计算。比如，Peterson 和 Barney 发现，他们的 76 个发音人在自然说话的时候，F_0 总是随着不同的元音而改变的。例如男人发音，高元音 [i] 的 F_0 平均是 136 赫，[ɪ] 是 135 赫，[ɛ] 130 赫，[æ] 127 赫，[ɑ] 124 赫，从 [i] 到 [ɑ] 一共相差 12 赫。从 [ɔ] 到 [ʊ] [u]，F_0 又逐渐上升，分别是：129 赫，137 赫，141 赫。但是研究

语言，仅仅从物理的观点来看问题是不够的，还应当把内在变化看成是一种心理现象，即感觉上的现象。比如对 F_0 随不同元音而变化的现象还要同时联系振幅的变化来观察。在自然说话的时候，低元音（比如 [ɑ]）的振幅比高元音（比如 [i]）的振幅要大得多。几十年前丹麦著名语言学家 O. Jesperson 曾经研究过这个问题。他做了一次实验，由他站在铁路的一根枕木上发不同的音，他的助手一边听一边沿着铁轨逐渐走远，看看什么音到距离多少根枕木远的地方会听不见。他们的发现是，低元音比高元音传得远。这就是一种内在变化。但是，我们平常说话的时候虽然有这种内在变化，而且这种变化可以相差好几个分贝，可是听起来好象是一样响。同样，根据 Peterson 和 Barney 的实验，不同元音的 F_0 的频率虽然是低元音比高元音低，可是听起来又好象音高是一样的。我本来以为，只有象英语那样音高不大重要的语言才存在这种现象，其实不然。象汉语这样有声调的语言也同样存在这种现象，而且程度和英语差不多。我曾经测量了很多普通话的音节，比如“米”和“马”，“米”中元音 [i] 的 F_0 频率要比“马”中 [a] 的高五、六赫；同样，比如“铺”和“怕”，“铺”中 [u] 的 F_0 频率也要比“怕”中 [a] 的高。但是，在听觉上这些音节都是一样高的。

但是，如果我们用电子计算机来进行综合语音的工作，把语音的 F_0 和 F_1 F_2 F_3 等等的各种值都输入进去，命令机器根据固定的振幅和固定的基频发出不同的音来，这时人们对这种语音的心理感觉就发生了变化。试比较：

自然说话	综合语音
db(低元音) ＞ db(高元音)	db＝db 响度(低元音)＜响度(高元音)
hz(低元音) ＜ hz(高元音)	hz＝hz 音高(低元音)＞音高(高元音)

这就是说，把振幅和基频完全固定下来，它们的响度和音高在感觉

上就会变得不同。此外，音长也有类似的现象，在自然说话的时候，低元音一般比高元音长，有时可以相差百分之十五到二十左右。比如，英语 bat 里的 a 比 bit 的 i 要长百分之十到二十，可是听起来象是一样长。由以上种种分析可以看出，语音作为物理现象和作为心理现象并不总是一致的。语言学家常常提到的“超音段成分”(suprasegmental)，也就是音强、音高和音长，都有它们的内在变化，对它们进行测量的结果，有时和心理上、听感上的印象不相一致，甚至刚好相反。

CVC 中的内在变化

内在变化不只限于元音本身，还包括元音及其前后辅音之间的相互影响。我们可以用 CVC 来代表“辅音＋元音＋辅音”这样的语音结构。在 CVC 里，后一辅音往往影响元音的音长，前一辅音往往影响元音的音高。

下面我们重点介绍辅音影响元音音长的问题。I. Lehiste 和 G. E. Peterson 十几年前合写了一篇文章，叫做《转接、音渡和复元音》(Transition, Glide and Diphthong)，比较详细地谈到了这个问题。比如英语 Bee [bi], bean [bin], bead [bid], beat [bit] 这些词，从声谱图上可以测量出，[bit] 的元音 [i] 比 [bid] 的要短得多，也许只有 [bid] 的元音四分之三长。[bin] 的 [i] 看去似乎最长，但是，除去后面应该属于 [n] 的共振峰的比较平行的一段，实际上跟 [bid] 的 [i] 差不多长。我们所要强调的是 [bid] 和 [bit] 当中的区别，也就是说，同是在舌尖音 [d] [t] 之前，元音 [i] 的长短不同。如果多测量一些例子，就会发现，同在清塞音之前，如 [bit]、[bik]、[bip], [i] 的长度相同；同在浊塞音之前，如 [bid]、[big]、[bib], [i] 的长度也相同。可见元音长短不同的原因不在后面辅音的发音部位，而在发音方法。从测量过的好几种语

言来看，都有这种变化。这就是一种内在变化。它是相当有规律的：一个元音如果后面没有辅音 (V#)，比较而言最长。后面有浊辅音(VÇ)，和没有辅音长度差不多。后面有清辅音 (VÇ)，元音的长度就比后面没有辅音和后面有浊辅音短得多。同是后面有清辅音，在清塞音前又要比在清擦音前短。那么，究竟是浊辅音把元音拉长了呢？还是清辅音把元音缩短了呢？从上面的分析已经知道，元音后面没有辅音 (V#) 和后面有浊辅音 (VÇ) 差不多是一样长的，可见是清辅音把元音缩短了。

英语里有两套不同的元音，一套比较长，一套比较短，传统语音学分别称为紧元音和松元音。比如 [i] 和 [ɪ] 这对元音，[i] 比较长，比较紧，[ɪ] 比较短，比较松。所谓松紧，一般认为是发音时肌肉比较松或紧，但这直到现在也没有找到证据。不管是长是短，是紧是松，清辅音使元音缩短的现象在这两套元音里都是一样存在的。在读松元音 [ɪ] 的词里，bit [bɪt] 里的 [ɪ] 要比在 bid [bɪd] 里短，bin [bɪn] 里的 [ɪ] 跟 bid 里的就差不多一样。

再举一个复杂一些的例子。比较 fate [feⁱt] 和 fade [feⁱd]、lope [loᵘp] 和 lobe [loᵘb] 这两对词，看看元音 [eⁱ] 和 [oᵘ] 的长短是不是也有变化。从声谱图可以看出，fate 发音时，由 [e] 渐渐变到 [i]，共振峰 F_1 由较高降到较低，然后舌尖抵住齿龈，堵住气流通路，这时有一段时间没有音；等到舌尖离开齿龈，气流冲出，就发出了 [t]。fade 发音时，同样由 [e] 到 [i]，然后舌尖抵住齿龈，堵住气流通路，但这时口腔里气压比较低，空气可以从声带当中漏过去，引起声带颤动，然后舌尖离开齿龈，气流冲出，发出 [d]。从元音共振峰的长度看，fade 的 [eⁱ] 比 fate 的 [eⁱ] 差不多要长一倍。lope 和 lobe 的情况也大致相同。[p] 是个清辅音，发音时声带不颤动；[b] 是个浊辅音，发音时虽然嘴唇闭着，

声带还是在颤动。lobe 的 [o^{u}] 比 lope 的 [o^{u}] 要长出四分之一到三分之一。

Lehiste 和 Peterson 还曾经对英语里的 [a^{i}] [ɔ^{i}] [a^{u}] 三个复元音做了实验，用下列三对词做例子：

tight [ta^{i}t]　　voice [vɔ^{i}s]　　lout [la^{u}t]

tide [ta^{i}d]　　noise [nɔ^{i}z]　　loud [la^{u}d]

从声谱图上看，[ta^{i}t] 在辅音破裂送气以后，共振峰就开始，但是弯曲得很厉害，在 [a] 的部分也许还不到 100 毫秒，就开始往 [i] 的部位移动，所以 F_1 迅速下降，F_2 迅速上升；然后到 [t]，先是舌尖闭塞，再破裂送气。[ta^{i}d] 发音的情况类似，但是 [a^{i}] 要长得多，因为到 [d] 的时候，舌尖虽然已经堵住了气流通道，但由于口腔里气压比较低，空气还是一喷一喷地进入口腔，声带颤动了大约六、七次，然后停止一二十毫秒，才发出 [d] 来。所以 [ta^{i}t] 里的复元音比较短，[ta^{i}d] 里的复元音要长出三分之一。同样，在另外两组例词中，[vɔ^{i}s] 里的 [ɔ^{i}] 比较短，[nɔ^{i}z] 里的 [ɔ^{i}] 比较长；[la^{u}t] 里的 [a^{u}] 比较短，[la^{u}d] 里的 [a^{u}] 比较长。

辅音影响元音音长的内在变化不限于元音和辅音直接相连的环境。有时候元音和辅音中间即使隔着一个别的音，如 [r] [m] [n]，也还是能发生这种内在变化，结果成为 $V\left|\begin{matrix}r\\m\\n\end{matrix}\right|C$ 这样的情况。比如，我们已经知道 said [sɛd] 的元音 [ɛ] 比 set [sɛt] 的长，因为前一个元音后面是浊辅音 [d]，后一个元音后面是清辅音 [t]。包含有同一元音 [ɛ] 的 [sɛnd] 和 [sɛnt] 情况也相同，虽然这时元音 [ɛ] 和辅音 [d] [t] 并不紧连着，中间隔着一个 [n]，但 [d] [t] 的影响仍然存在，[sɛnd] 里的 [ɛ] 比 [sɛnt] 里的 [ɛ] 还是长得多。比较 hurt [hɛrt] 和 herd [hɛrd]，也是如此，[hɛrt] 的 [ɛ] 要短得多，只是中间隔着的不是 [n]，而是 [r]。但 herd

里的 [ɛ] 和 burn [bɛrn] 里的 [ɛ] 就差不多一样长，因为它们后面都是浊辅音。

对 CVC 的内在变化研究得最系统、最有成绩的是伊利诺(Illinois) 大学的 A. S. House 和 G. Fairbanks。Fairbanks 已经故去。House 是他的学生。他们合写过一篇文章叫《辅音环境对元音次要声学特征的影响》(The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels)。文章介绍了他们对 [i] [e] [æ] [ɑ] [o] [u] 六个元音在内在变化中音长、音高、音强方面的变化进行的系统的测量，所得的数值列成下表:

	[i]	[e]	[æ]	[ɑ]	[o]	[u]
				音　长(秒)		
总计 (12)	0.199	0.225	0.244	0.236	0.221	0.195
清塞音(3)	0.138	0.171	0.184	0.180	0.157	0.138
清擦音(2)	0.177	0.199	0.215	0.218	0.187	0.161
鼻　音(2)	0.209	0.238	0.253	0.235	0.241	0.217
浊塞音(3)	0.215	0.251	0.276	0.267	0.244	0.215
浊擦音(2)	0.277	0.283	0.304	0.295	0.293	0.261
			基频(周/秒)			
总计 (12)	127.86	123.03	119.78	118.00	124.64	129.83
清塞音(3)	132.07	126.44	122.00	120.97	128.57	133.79
清擦音(2)	129.59	124.05	119.80	119.59	125.86	132.86
鼻　音(2)	125.89	119.93	118.75	117.78	122.64	129.83
浊塞音(3)	125.34	121.02	117.87	115.80	123.13	125.26
浊擦音(2)	125.58	123.00	119.69	116.25	121.22	127.69
			相对强度			
总计 (12)	12.43	14.94	12.35	11.97	15.49	14.94
清塞音(3)	4.16	8.30	5.50	5.78	7.35	4.91
清擦音(2)	12.43	12.70	11.39	11.17	12.83	11.48
鼻　音(2)	17.05	15.96	17.18	12.14	16.84	17.97
浊塞音(3)	13.80	16.65	13.66	14.94	18.29	14.78
浊擦音(2)	18.19	23.58	16.74	17.42	24.82	30.67

他们为 [i] [e] [æ] [ɑ] [o] [u] 六个元音分别加上十二个不同的辅音。表的上部是元音音长的比较。可以看出高元音短，低元音长。[i], [u] 是高元音，最短，[i] 是 199 毫秒，[u] 是 195 毫秒；[æ] 和 [ɑ] 是低元音，分别是 244 毫秒和 236 毫秒。[æ] 和 [i] 相差达 45 毫秒。每个元音处在不同的辅音之前又会改变它的长度。比如 [i]，在清塞音之前长 138 毫秒，在浊擦音之前长 277 毫秒，后者差不多是前者的一倍长。

表的中部是元音音高(即 F_0)的比较。这对声调研究很有用。因为通常所说的声调调值，比如汉语普通话阴平 55、阳平 35 等等，只是一种很简单的说法。实际上元音舌位越高，F_0 也越高；元音舌位越低，F_0 也越低。所以，[i] 是 127.86 赫，[u] 是 129.83 赫，[æ] 是 119.78 赫。[i] 比 [æ] 约高 8 赫，[u] 更高出约 10 赫。与此同时，元音出现在不同的辅音环境中，在音高方面也有它的内在变化：在清辅音后，F_0 就高一些，在浊辅音后，F_0 就低一些。比如汉语普通话“骂” [ma] 的 F_0 如果用电子计算机进行计算，和“怕”的调型就不同，这是因为 [m] 是个浊辅音，而浊辅音是会把 F_0 往下拉的。(参看 15 页图 3-2) “骂”和“怕”曲线的形状不仅最高点有很大不同，终点也有相当的区别。类似的现象在非声调语言如英语里也是常见的。例如 pete 和 beat 的调型也不一样，beat 的 F_0 开头是比较低的。

表的下部是元音振幅的比较。总的说来，在清辅音的环境里，元音的振幅比较低；在浊辅音的环境里，振幅比较高。至于单个元音的情况，前面已经谈过，这里就不提了。

还有这样一个问题：在 C_1VC_2 里，为什么 C_1 影响 V 的音高，C_2 影响 V 的音长？这个问题很难回答。目前虽然有很多解释，但是没有一种能被所有人接受。下面举出一种解释，是说明 C_1 的清浊不同为什么会影响 V 的音高的。上面已经提到，浊辅音

会把后面元音的 F_0 往下拉。可以这样解释这种现象：发清辅音例如 [p] 的时候，元音和辅音的破裂同时开始，气压一开始就比较大。如果别的条件相同，气压越大，声带颤动的速度就越快，所以一开始 F_0 就很高。发浊辅音例如 [b] 的时候，情况就不同，嘴唇还没有张开时声带就颤动了。而声带每颤动一次，就要漏出一些气来，气腔里的气压也就增高一些，颤动的次数多了，气压就有可能达到平衡。而为了避免气压平衡，就把声带、喉头往下降，使气腔扩大。喉头一下降，声带就比较松，颤动也就比较慢，F_0 自然也就下降一些。发浊音的时候喉头往下降，这是完全可以测量出来的。

语音变化和音渡

现在谈谈元音和辅音之外的问题。近二十几年来发现，在 CVCVCVCV 这样一串东西里，元音和辅音之间会产生很多不同的现象，以至于单用元音和辅音来代表这一串声音有时会感到不够。比如英语的 gray day [gre^ide^i] 和 Grade A [gre^ide^i]，用国际音标标音是相同的，可是说英语的人听起来和说起来完全不一样。这个问题在五十年代和六十年代曾经引起美国语言学界的广泛讨论。当时为这种现象提出了内部音渡 (internal juncture) 这一名称。一般的说法是，只用元音和辅音串在一起来代表语音的成分是不够的，因为有时候词和词之间的音渡会发生不同的情况。所以，从音位学观点看，可以借助音渡把 gray day 标写成 /grei+dei/，把 Grade A 标写成 /greid+ei/。音渡现在一般都是用符号 / + / 来代表的。

这样的例子在汉语普通话里大概不容易找到，因为普通话音节结构都是相当简单的，音节末尾的辅音种类很少。英语的例子就很多，I. Lehiste 收集了大约一百多对这样的例子，如：

gray day	（阴天）	Grade A	（一级品）
white shoes	（白鞋）	why choose	（为什么选）
see Mabel	（看梅博）	seem able	（似乎能）
ice cream	（冰激凌）	I scream	（我喊叫）

Lehiste 在写她的博士论文《内部开音渡的声学语音研究》(An Acoustic-phonetic Study of Internal Open Junture) 时对这些音渡做了大量的实验，目的是要看看在声谱图上是否有不同的表现。结果发现，同一个音位上的单位，同一个音渡，在各种不同的元音和辅音的环境中所产生的现象，是不同的。比如在 white shoes 和 why choose 中，两个都是 [hwaⁱtʃuz]，但是前一个的 [ʃ] 长，后一个的短得多，音渡在这里表现为擦音的长短。在 see Mable 和 seem able 里，两个都是 [simeⁱbl]，但是在由一种叫做连续振幅显示 (continuous amplitude display) 的窄带声谱仪所做的图上可以看出，see Mable 中 [m] 后面的元音 [eⁱ] 相当完全，F_0 相当高，seem able 的 [eⁱ] F_0 就很低，音渡在这里表现为元音 F_0 频率的高低。gray day 和 Grade A 都是 [greⁱdeⁱ]，情况跟 see Mabel 和 seem able 相同，只是没有那样显著。ice cream 和 I scream 都是 [aⁱskrim]，但后一个的 [k] 送气程度不如前一个强，音渡在这里表现为清塞音的送气程度的强弱，这是用耳朵也可以听出来的。

Lehiste 用仪器来测量对比，避免了耳听时可能产生的主观成分。她的测量说明，存在着两种不同的音位：通常元音辅音所用的符号（如 / e /）是一种音位，音渡这样的符号 / + / 又是一种音位。这两种音位有着根本的区别。前一种是一般的音位，每个音位都有它们各自的特征。/ + / 这样的音位并没有自己固定的特征，它在语音上最重要的功能是使与它相邻的音受影响从而可以区别不同的词。受到什么样的影响，要看在它附近的是什么样

的音位。比如，如果是元音，它会降低后面元音 F_0 的频率；如果是擦音，它会使后面的擦音增长；如果是清塞音，它会改变后面清塞音的送气程度。

七　语音的图谱

语图和语图仪

我们说话发出的每一个声音，都可以用仪器把它画成图谱。这种仪器叫“声谱仪”(Spectrograph，也叫 Sound Spectrograph)，是贝尔公司在第二次世界大战期间秘密发明的，战后才公开。用声谱仪画出来的图叫“声谱图”(Spectrogram)。最早出售声谱仪的是凯氏电气公司(Kay Elemetric Corp.)。给它起了一个商业的名字，叫做 Kay-Sonograph，译成“语图仪”。画出来的声谱图就叫 Sonogram，译成“语图”。现在，除了凯氏电气公司以外，还有一些小公司也开始做不同种类的声谱仪，但是在世界上用得最多的还是凯氏的语图仪。

做这方面实验本来是为了发明一种能使聋人看见语音的机器。人们一边说话，一边就能在电视上把它显示出来，聋人就可以看着电视判断说的是什么，所以叫做“可见语言”(visible speech)。现在，分析语音的图谱已经成为语音学很重要的内容。

附页图 7-1 是英语“speech production”的声谱图。[①] 图的上部是宽带 (broad band) 声谱，下部是窄带 (narrow band) 声谱。我们说话时声带不断颤动，声门每一次打开，喷出气来，在咽腔、口腔、鼻腔引起共振，就有一个谐波。图中的宽带声谱图上部把不同的图样标成 A, B, C, D 四种，其中 B 类都有许多条很清楚的直

① 图取自 W. S-Y. Wang and R.S. Tikofsky, 1960, «Speech», McGraw-Hill Encyclopedia of Science and Technology, 593—9。

线，每一条直线就代表声带颤动一次。因为线和线的距离很近，谐波的高峰连起来，在图上形成横条，就是这个声音的共振峰。由下而上，第一横条是 F_1，第二横条是 F_2，第三横条是 F_3。我们也可以看出 F_0，那就是两条直线之间的距离。

声谱图的图样

下面想比较系统地介绍一下声谱图的各种不同图样和它的术语，然后再分析几个不同的声谱图。

1．间隙 (gap) *G*

从宽带的声谱图上，可以看到六种不同的基本图样。最简单的就是“间隙”(gap)，用 *G* 来代表。发塞音的时候，尤其是清塞音，在破裂之前有一段时间并不发出任何声音。图 7–2 是“他病了”的声谱图，在“他”和“病”之间至少有一百毫秒的时间没有声音，那一段时间就是间隙 *G*。附页图 7–1 中的 D 也都是间隙 *G*。

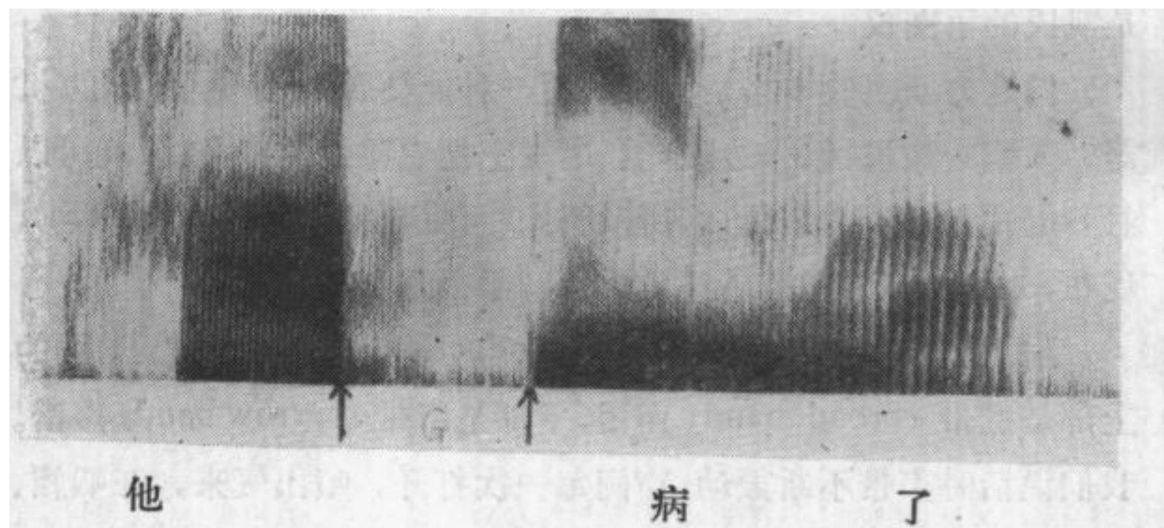


图 7-2

2．宽横杠 (voiced bar) *V*

清塞音前头的间歇一点声音也没有，浊塞音则不同。附页图 7–3 是 [ba] 的声谱图，可以看到在间隙的最下面有一条比较宽的

横杠，说明声带是颤动的，是有声音发出的。我们管这横杠叫“宽横杠”(voiced bar)，用 V 来表示。

3．冲直条 (spike) P

发塞音的时候，例如 [p]，嘴唇一张开，就发出音来，在声谱图上就出现一条直线。这条直线叫“冲直条”(spike)，用 P 来代表。附页图 7–1 中的A都是冲直条。从图 7–2 “病”的开头也可以看出冲直条来。

4．共振峰纹样 (formant pattern) F

关于共振峰，前面已经谈过好多次了。在声谱图上表现为一条条的横杠，叫做“共振峰纹样”(formant pattern)，用 F 来代表。从下往上，分别叫做 F_1，F_2，F_3，等等。附页图 7–1 中的B都是共振峰纹样。

5．杂乱纹样 (noise pattern) N

共振峰纹样的声音起源于声带，杂乱纹样的声音一般说来起源不在声门，而在口腔的某个部位。清擦音 [ɸ，θ，s，ʃ，ɕ，x，h] 等在口腔里的发音部位虽然不同，在声谱图上看到的都是“杂乱纹样”，用 N 来代表。附页图 7–1 中的C都是杂乱纹样。

6．双有纹样 F/N

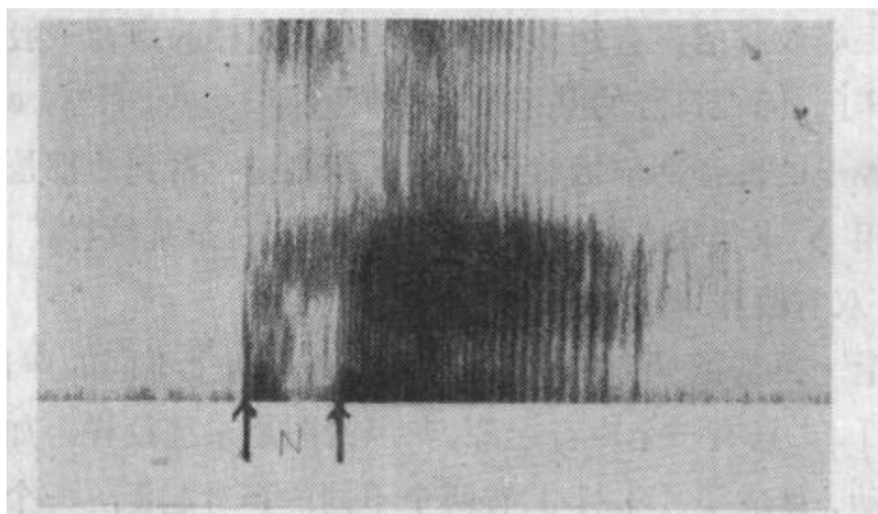
最后一类是把 F 和 N 合起来，具有两个不同的声音起源。比如发 [s]，只有一个声音起源，在声谱图是杂乱纹样。如果声带同时颤动，就成了 [z]。[z] 有两个不同的声音起源，一个在上门齿，一个在声带。从上门齿来说是 N，从声带来说就是 F，把这两个合起来，就是 F/N，所以管它叫做“双有纹样”。

宽带的声谱图，就只有这六个基本图样。要想知道所看到的究竟是什么音，那就要仔细看它的共振峰纹样是怎样的，杂乱纹样是怎样的，等等。

最近七、八年，在一些文献里有时还会遇到一些别的术语。

VOT 就是很常见的一个，现在有很多人对它感兴趣。VOT 是"voice onset time"的缩写，也许可以把它译成"相对清浊度"。

VOT 是专对着塞音说的。比方我们说 [ba]，声带颤动得很早，从附页图 7-3 [ba] 的声谱图上可以看出，在 [b] 的冲直条 *P* 之前，也就是在嘴唇打开以前很久，就已经有了宽横杠。所谓 VOT，就是这个宽横杠所用的时间，如果是 110 毫秒，这个音节的 VOT 就等于 −110。在图 7-2"他病了"的声谱图上，"病"的前面就没有那个宽横杠，冲直条 *P* 之后马上就有共振峰纹样 *F*，在这种情况下，VOT 就等于 0。图 7-4 是普通话"怕" [p'a] 的声谱图。"怕"是一个送气的清塞音，嘴唇张开后，可能有 40—50 毫秒时间是杂乱纹样 *N*，然后才有比较清楚的共振峰 *F*。*N* 的时间如果是 50 毫秒，那就是说，VOT 等于 +50。概括地讲，VOT 就是冲直条 *P* 跟宽横杠 *V* 或共振峰纹样 *F* 之间的时间关系。



怕 [p'a]

图 7-4

对 VOT 已经做过很多实验。有的实验发现，刚生下来不到一个月的婴儿听 VOT 的规律的方法，和大人差不多完全一样。这是人类对语言有先天本能的一个相当具体的证明。最近有不少人在研究和讨论 VOT。

以上六种基本图样可以描写、叙述一切声谱图。下面把这六种图样和语音的类别加以对比。为了便于叙述，我想还是用传统的方法来给语音分类比较方便。最基本的分类就是这个音是不是响音 (sonorant)。响音可以分成几大类，最大的一类就是元音，元音的基本图样就是共振峰纹样 F。

响音中的辅音可以分成流音 (liquid) 和鼻音两种。流音根据它是否颤动又可以分成两类。不颤动的流音如英语 l r 等都是共振峰纹样 F。r 和 l 最明显的区别在 F_3 的位置。例如 seer 里的 r，F_3 是跟着 F_2 一起下降的，而 feel 里的 l，F_3 是往上升的。

有的流音是颤动的，是颤音 (trill)。有三种不同的颤音：[ɾ] [r] [ʀ]。[ɾ] 是舌尖打在硬腭上，只打一次，日语的 [ɾ] 就是如此。只颤动一次，就只有一个冲直条 P。[r]，俄语的要颤动六、七次，意大利语和德语有些方言要颤动好多次，就等于有好多不同的冲直条 P:$PPPPP$…。小舌颤音 [ʀ] 也有好些条冲直条 P，不过频率上有区别。（颤音颤动的次数有时能区别不同的词，南美一些西班牙方言的颤音，颤一下和颤几下就是不同的词：deros, dellos 的分别就是如此。南斯拉夫有些方言也有这种区别。）

另一套响音是鼻音。鼻音的基本纹样也都是共振峰纹样 F。鼻音的声音从鼻腔出来，口腔里阻碍的位置不同，声波就会受到影响，所以 [m] [n] [ŋ] 的共振峰频率有区别，听起来也不一样。但是这个区别是非常小的。如果听人发 [m]]n] [ŋ] 看着嘴，可以比较清楚地听出来发的是哪个鼻音；如果发音人背过身去发音，就不大容易区别了。由此可以看出，至少从鼻音来讲，单靠它本身的共振峰频率是不够的，一定还要依靠别的东西才能够知道发出来的是哪个鼻音。这个问题后面还会谈到。

响音之外就是阻塞音 (obstruent)。可以分成两大类。一类是擦音。清擦音在声谱图上的表现是杂乱纹样 N，浊擦音是双有纹样

F/N。Peter Strevens 对擦音非常有研究，他在 1960 年写了一篇《人类言语摩擦噪音的频谱》(Spectra of Fricative Noise in Human Speech)，其中有这样一个表格:

	一	二	三	
	Φ, f, θ	s, ʃ, ç	x, χ	h
频率范围	大	小	中	
频率部位	低	高	中	
幅 度	低(弱)	高(强)	中	

Strevens 一共用了九个清擦音进行研究，没有包括中国很常见的卷舌音 [ʂ]。他对这九个音进行了细致的分析，得出的结论可以分三方面来说：一是频率范围，二是频率部位，三是幅度。他把这九个清擦音分成三组。（把 [h] 放在第三组是不适当的，因为其余八个清擦音的发音部位是固定的，[h] 不是这样，它在什么元音之前，部位就要受到这个元音的影响，在声谱图上也就象发这个元音的部位，所以不应该列在表中第三组里。）

频率范围是第一组最大，[f] 一直可以降到两千多，高到六千多，范围最大。第三组 [x,χ] 是中等，第二组 [s,ʃ,ç] 范围最小。

频率部位是第二组最高，因为它的频率高。第一组最低，因为它的频率最低。第三组中等。

从幅度看，第二组强度最高，第一组强度最低，第三组中等。为什么 [s,ʃ,ç] 的幅度最高呢？这和发音时舌头平面的形状有关系。发第一组和第三组的时候，舌面是相当平的。发第二组 [s, ʃ, ç] 的时候，舌头当中是凹下去的，声门送出来的气因此就比较集中，气流自然比较强，声音也就比较大。

另外一类阻塞音可以分成塞音 (stops 或 plosives) 和塞擦音 (affricates) 两类。在声谱图上最大的特征就是有间隙 G（如果是

浊的，同时有宽横杠 V)，在间隙后面有一个冲直条 P。如果是送气的，再加一个杂乱纹样 N。

转接和音轨

以上所谈的主要是仪器分析的结果。有时候单靠仪器分析是不够的，还要做听觉的实验，从听觉得出答案。下面介绍两个剪磁带的听觉实验。这两个实验是针对一个音段本身的信息和它附近的音段有冲突的时候听觉怎样而做的。一个是 K. Harris 做的，专门研究擦音，写成了《摩擦辅音的一些声学音征》(Some Acoustic Cues for the Fricative Consonants) 这篇文章。另一个是我在二十年前做的，专讲塞音，写成了《转接和除阻作为末尾塞音的感知音征》(Transation and Release as Perceptual Cues for Final Plosives) 这篇文章。

下面先谈擦音。比方 sa 这个音节，在声谱图上是分成两段的：一段是杂乱纹样 N，另一段是共振峰纹样 F。从图上可以看出来，s 有两种不同信息：一种是 s 本身，从杂乱纹样 N 的频率和幅度的关系表现出来。另一种是在 s 和 a 交界的地方。我们单独读 a，就只是三个平行的共振峰，是很简单的；如果前头有 s，F_1，F_2，F_3 的开头部分就都变了，有的朝上，有的朝下，这是杂乱纹样 N 对共振峰的影响。s 的第二种信息就在这方面，也就是在"转接" (transition) 上。转接就是一个辅音对与它邻近的元音的共振峰频率的影响。不同的辅音对共振峰有不同的影响，表现出来的转接也不同。

Harris 是差不多和我同时做的实验。我们并没有互相通信，但是得出的结果相当一致。Harris 把很多不同的擦音和很多不同的元音结合在一起，录在磁带上，然后用剪刀在擦音和元音交界的地方剪开，让转接留在元音上头。比方 sa，si，fa，fi，辅音同

样也要受元音的影响，剪开就是这样：

s_a / $_s$a　　f_a / $_f$a

s_i / $_s$i　　f_i / $_f$i

然后把它重接起来，接成这个样子：

s_a $_f$i

这样的音节人是发不出来的。把很多这样用磁带剪接出来的音让人听。象上面接的 $s_a$$_f$i，一般人听了都说这个音节是 si。可是，如果接成这个样子：

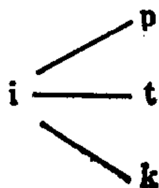
f_i $_s$a

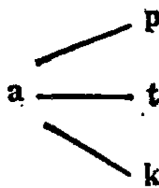
听到的人就都说是 sa，而不是 fa 了。

Harris 在这个实验里的研究结果和前面提到的 Strevens 的研究结果密切相关。在转接的力量和声谱中幅度的力量有矛盾的时候，最主要的是幅度的强弱。如果前面这个擦音幅度比转接强，听到的是这个擦音的声音；如果前面这个擦音幅度比转接弱，听到的就是由转接带出来的声音。

一个声波里有成千上万的信息，要了解人怎样听懂语音，就要了解这些信息中哪些重要，哪些不重要。怎样做到这点呢？可以把不同的信息综合起来，看看人的听觉反应如何。Harris 的实验就是这样做的。

我做的塞音试验和 Harris 的结果差不多。我用的是几个不同元音，后头是 -p，-t，-k 和 -b，-d，-g。例如：





有时候转接告诉听话人后头是 t，但实际上后头跟着的不是t，而是k。（例如把 at 的转接后面的 t 剪掉，换上 k。）这种实验至

少有两点值得注意。一点是清塞音的除阻比浊塞音强得多。如果除阻是清塞音，听去往往就是这个音，如果是浊塞音，听去往往就是转接那个音。另一点是转接的力量由它弯的角度决定。如果弯得厉害，这个转接在听觉上就很有力量；如果相当平，就不大有力量，听的时候，就会使人注意到除阻。

这两个实验都专门指出转接的重要性。前面曾经谈到，鼻音 [m] [n] [ŋ] 本身在声谱图上不好分别。在这种情况下，转接就非常重要。

共振峰的转接既然如此重要，是不是能找到它的一些规律呢？K. Potter, A. Kopp, C. Green 合写的《可见语言》(Visible Speech) 这本书里做了不少声谱图，其中的转接很清楚。Potter 他们把转接开始的地方叫"音轴" (hub)。可是始终找不到音轴和语音之间的固定关系。不同的音轴有时候听起来是同一个声音，有时候听起来是不同的声音。声谱图里的东西也许太复杂了，所以只从图上看不出其中的规律。如果用简单的语音合成仪器把 [b] [d] [g] 等等合成出来，再有计划地把其中的转接一一剪接，看人的耳朵反应怎样，就有可能找到其中的规律。C. Delattre, M. Liberman, S. Cooper 合写的《辅音的音轨和转接音征》(Acoustic Loci and Transitional Cues for Consonants) 这篇文章在这方面做出了很大贡献。

Delattre 他们所用的语音合成仪器非常简单，叫做"模式播放器" (pattern playback)。它的作用和语音分析完全相反。先画一张声谱图，送进仪器，就能把它合成，播放出声音来。送进去的是画的声谱模式，播放出来的是声波，所以叫做"模式播放"。送进去的声谱模式如下页图 7–5 的三个图。第一个图送进去听起来前头有一个 [b]，第二个听起来象有 [d]，第三个象有 [g]。三个图所有的 F_1 的转接都差不多，都是挺直地往下指。（只要是塞音，F_1 都

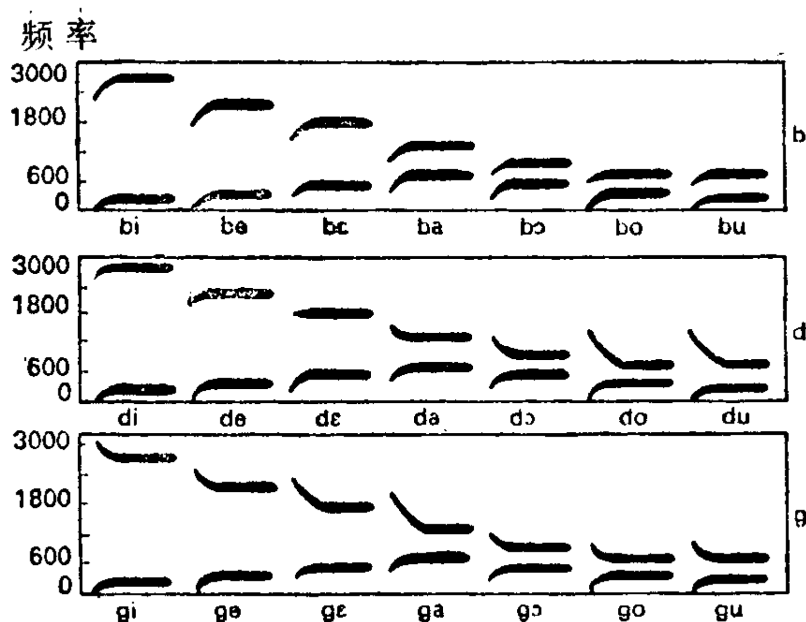


图 7-5

是往下指的。) F_2 就不同了，转接随着辅音和元音当中的关系而变化。[b] 的 F_2 虽然都往下指，但是 [bu] 只往下指一点，[bi] 就往下指得很多。[d] 的 F_2 更特殊了，有的往下指，有的往上指，有的是直的。[g] 的 F_2 都是往上指，可是程度也不一样。

我们已经知道，元音是因为口腔形状不同才有不同共振峰的。转接也是这样。口腔形状不同对转接的影响也不一样，但同样的形状应该有同样的影响。Delattre 他们从这个概念出发，从音轴进一步研究到音轨 (locus)。图 7–6 可以说明音轨这个概念:

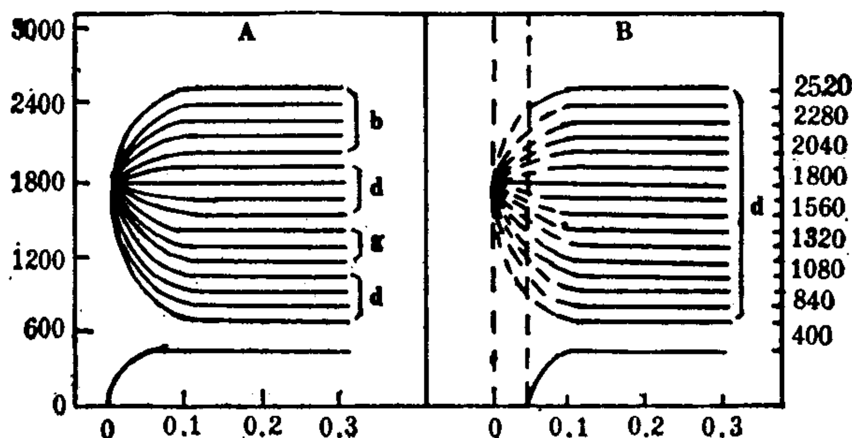


图 7-6

音轴是从发音才开始的，音轨是假定在发音之前有一段时间舌头和嘴在动，但是还没有发出音来。图 7–6 的 A 图和图 7–5 那三个图是一样的，只不过把它们画在一张图上，不同的 F_2 跟 F_1 相配，听起来有时是 [b]，有时是 [d]，有时是 [g]。B 图把不发声的这部分去掉，大约去掉 50 毫秒的样子(图中两虚线之间)，结果所有 F_2 加上 F_1 都成为 [d] 了，只是元音有时是 [i]，有时是 [a]，有时是 [u]，但前头辅音听起来都是 [d]。有的时候，F_2 根本不动，但是 F_1 在动，加起来也还是一个很清楚的 [d]。

由以上得出的结论就是：发音部位不同，音轨也就不同。[d] 是舌尖音，有它固定的音轨，就是 [d] 的不弯的 F_2 直指着的频率，在 1,800 赫左右。[b] [p] [m] 这些唇音的音轨是在 700 赫左右，[d] [t] [n] 这些舌尖音的音轨是在 1,800 赫左右。这样就找出了规律。唇音和舌尖音是由嘴唇和舌尖发出的，和元音的舌位关系不大。[g] [k] 是舌根音，和元音舌位的关系很密切。因此，一般讲来，[g] [k] 这类舌根音有两个不同的音轨。象 key [kiː] 开头的辅音相当靠前，是 [c]，could [kud] 开头的辅音相当靠后，是 [k]，二者音轨也不同，一个在 1,200 多赫，一个在 3,000 多赫。这是因为它们的发音部位和它后头的元音关系太密切了。

前面已经提到，一个声波有很多信息，有的信息和我们的听觉有关，有的和听觉无关。我们要把这两种信息区别开来，有用的那部分信息就叫做“音征”(cue)。比方 [b] 的音征就是在 700 赫左右的音轨。

下面根据我们刚刚学到的知识试着分析一张声谱图（见附页图 7–7)。这是一张早期的声谱图，频率范围只有 0—3,500 赫，3,500 赫以上的频率就看不见了。(现在做出来的图一般可以到 8,000 赫，甚至 16,000 赫。）先看第一个共振峰纹样，F_1 非常低，F_2 非常高，从这两个共振峰的关系可以确定它是一个 [i]。

[i] 的前头是杂乱纹样，前面没有冲直条，下面没有宽横杠，可以肯定是个清擦音。在 [f] [θ] [s] [ʃ] [h] 这些清擦音中，[h] 比较特殊。它的幅度比较低，在口腔里的部位不固定，最不影响元音的共振峰。和后面 [i] 的共振峰连起来看，可以肯定 [i] 的前头是个 [h]。

[i] 的后面又是杂乱纹样，幅度好象很强，比 [h] 强得多；然后看它的转接，F_2 的转接大致在 1,500 到 2,000 之间，这是舌尖音的音轨，所以可以肯定它是个 [s]。（这个 [s] 不很典型，典型的[s]的范围要高一些。）

[s] 之后又是共振峰纹样，肯定是复元音，因为它的共振峰动得很厉害，从比较低的元音部位变到比较高，而且一开始就相当高。我们还可以肯定这不是后元音，因为后元音的 F_2 是相当低的。所以，这是一个开始相当高，后来更高的复合前元音，也就是 [ei]。

把这个 [ei] 和图中最后的共振峰纹样对比一下，可以肯定最后也是一个复元音。两个复元音的共振峰收尾部位差不多，都是[i]。但是后一个在 [i] 的前面是一个后元音，而且相当高，可以肯定它是 [ɔi]。

[ei] 之后的图样比较难分析。先是一个浊擦音，后面的音轨降得很低，所以是个唇音，也就是 [v]。[v] 之后先是间隙，然后是一个冲直条，间隙之下有宽横杠，肯定是浊塞音，从它后面的音轨可以看出是舌尖音 [d]。[d] 的后头又是个元音，很短，越短越难分析。不过英语里元音读轻音的时候，音质常常会出现 schwa（非重读央元音，即 [ə]），从共振峰部位，可以断定这个很短的元音就是 schwa [ə]。[ə] 的后头又有一个间隙和冲直条，是浊塞音，它后面元音的音轨很低，因此是个唇音 [b]。

把以上分析综合起来，全句就是 [hiseivdəbɔi]，也就是 “He saved a boy”（他救了一个男孩子）。

要把声谱图的图样解读出来，是很难的事，可以说是一种艺术，要有丰富的经验才行。首先必须知道声谱图上是什么语言，预先了解音位之间的关系，然后才有可能看懂图谱。

窄带声谱图

下面分析一下窄带声谱图。先看附页图 7-1 里的下图，那就是窄带声谱图，内容仍旧是 “Speech production”。窄带声谱图和前面所谈的宽带声谱图很不一样，最显著的不同是有一条条横线。同一个声谱仪，插进不同的滤波器，做出的图样就不同。宽带声谱图常用的滤波器是 300 赫，窄带声谱图常用的是 50 赫。它们的不同作用可以用频率对幅度图来说明：

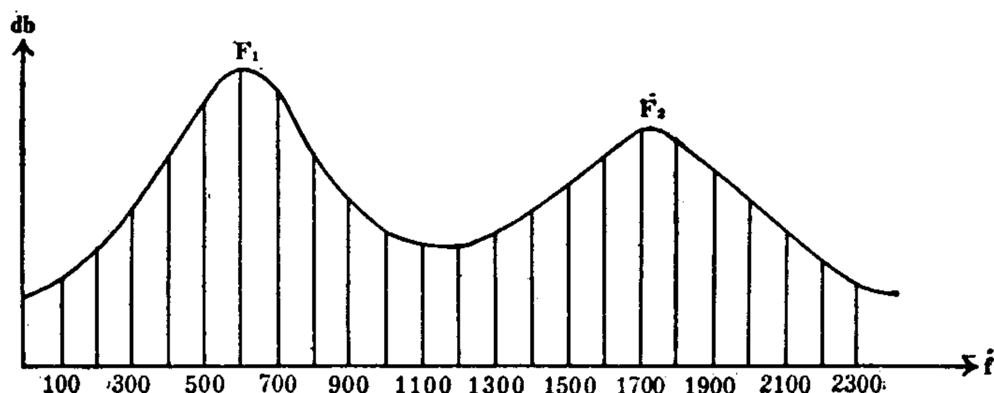


图 7-8

声带发出来的是一套谐波，通过口腔和咽腔，有了它的共振峰 F_1，F_2，谐波之间的距离就是它的 F_0。如果 F_0 是 100，下面的谐波就是 200，300，400，…图 7-8 的第一个共振峰 600 赫，第二个共振峰和谐波的频率并不一致，大约在 1,750 赫。

宽带滤波器的带宽是 300 赫。声谱仪动的时候，滤波器随着移动改变它的频率。滤波器到哪里，那里就画出那个范围内的电压。电压越高，画得越黑，画成一个时间对频率的图。不管滤波器移动到什么部位，它的范围总是包括三个 100 赫的谐波，画出来的只是

一个广泛的共振峰频率，不能把每一个各别谐波的频率画出来。

窄带滤波器的带宽是 50 赫，只是 100 赫谐波的一半。在它移动的时候，在 100 赫谐波之前没有电压，到了第一个谐波，有了电压，就画出一横。因为它的带宽只有 50 赫，所以每一条 100 赫的谐波都能画出来，成了一条条的横线，这样就可以直接看到每一条谐波。如果横线往上走，那就是 F_0 的频率在上升，也就是音高在升高，例如图中的 “speech” 就不是 [→spitʃ]，而是 [↗spitʃ]。我们已经知道，F_0 的频率和第一个谐波是相同的。要想知道 F_0 频率的变化，按说应该测量第一个谐波的变化。但是它离基线太近，变化不明显，因此一般都用第十个谐波来测量，因为第十个谐波的变化比第一个要大十倍，变化明显得多。比如第一谐波从 100 赫变到 110 赫，只变动了 10 赫；到了第十个谐波，就是从 1,000 赫变到 1,100 赫，变动了100赫，测量 100 赫自然比 10 赫容易得多。

在宽带声谱图上可以很清楚地看出共振峰，虽然也能看出F_0，但很不清楚。在窄带声谱图上可以看出共振峰，但是很模糊，清楚的倒是 F_0。两种声谱图，各有它的用处。研究声调和语调，就必须用窄带声谱图。

电子计算机声谱图和腭图

上面谈的都是用声谱仪做出的声谱图。近几年来国外一些语音实验室已经不常用声谱仪了，因为有些问题用电子计算机计算更方便，更准确。下面是我们的实验室专门为我们这次讲座用中型电子计算机 PDP 11 给英语“boy”画的两个图。（见下页图 7–9 和图 7–10）图做得并不很好，但是可以从中了解一些情况。

从图 7–9 可以看出，冲直条开始后有五个共振峰。F_1 比较低，差不多 400 或 500 赫，然后是 F_2，F_3，F_4，F_5。前面说过，共振峰的带宽跟它的频率有关，频率越高，带宽就越大，画出来的点

子也就越多。所以 F_1、F_2 是相当窄的线，到 F_5 就很宽了。从图上还可以看到 F_2 动得很厉害。横轴的刻度是100毫秒。在第一个100毫秒，F_2 就升得非常快，是从后低元音 [ɔ] 到高元音[i]。到200毫秒以后，F_1 的频率就不大靠得住了。这时声带颤动非常慢，到350毫秒，发音就停止了。

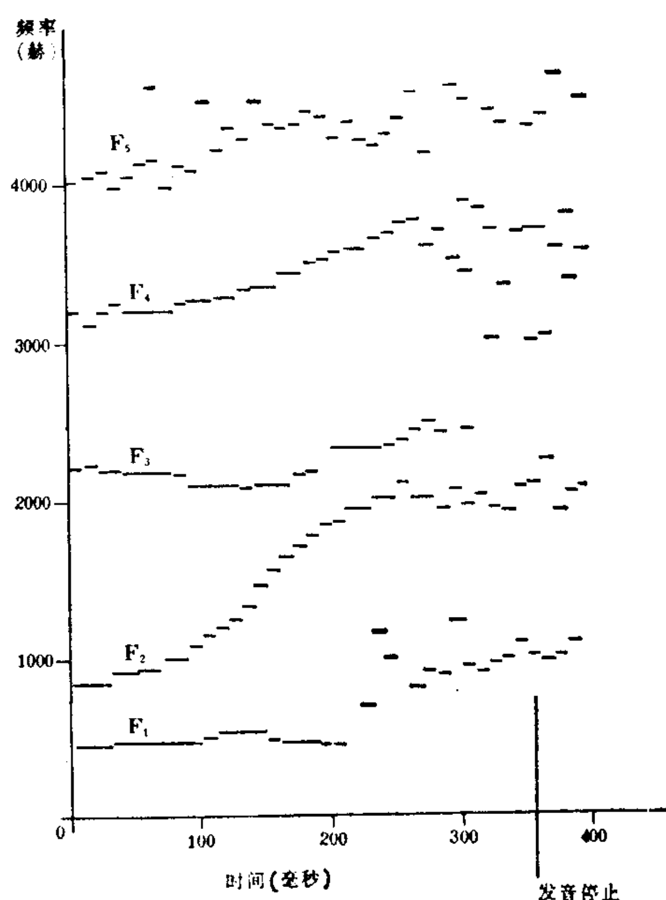


图 7-9

图 7-10 是 "boy" 的立体频谱图。一共有三个方面：横轴是频率，纵轴是分贝，斜线是时间。在图上一共有

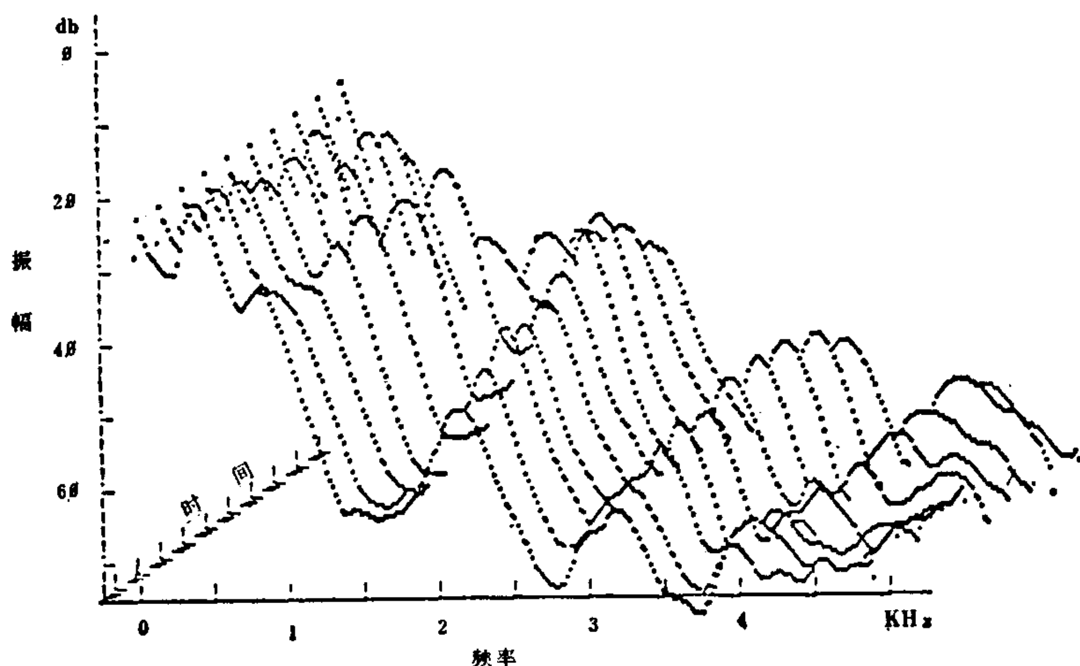


图 7-10

十个频谱。这十个频谱是从冲直条开始后的最前面 256 毫秒中取出来的，每 25.6 毫秒画出一个频谱，一共画出十个，每个频谱之间的距离就是 25.6 毫秒。计算机可以有伸缩性，相隔的时间不一定是 25.6 毫秒，可以更长一些或更短一些，也可以画一个平均数等等，不象声谱仪的作用那样固定。现在做声学方面的语音分析差不多都用这个方法。

最后谈一谈腭图 (Palatagraphy)。

前面曾经说到发 s 和 ʃ 的时候舌面中间是凹下去的，当中有条沟。这怎么能够证明呢？传统的办法是用假腭。但是这办法非常麻烦，而且结果也不大可靠。近年来对假腭有不少改进。例如发音后先不把假腭拿出来，而是伸进一个镜子去，对着镜子照相，也可以得到一张腭图（见附页图 7-11）。最近加州大学用电子计算机做成一种电子假腭，画出电子腭图，象下面这个样子：

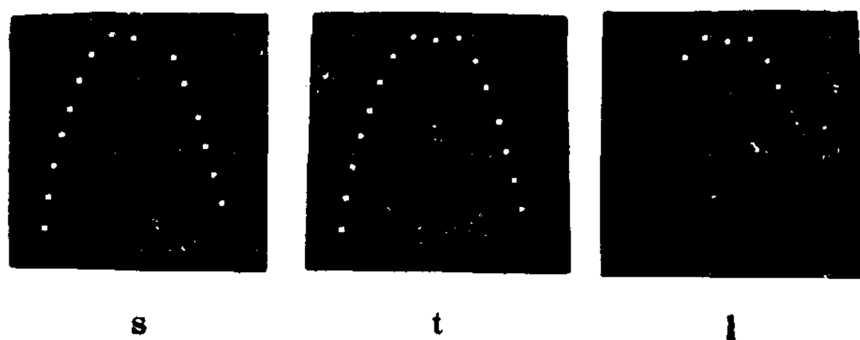


图 7-12

这种假腭有几十个电极，舌头一碰到电极，就会通电。电子计算机根据通电的情况进行计算。结果显示在荧光屏上，用照相机照下来，就是一张腭图。这三张图只是初步实验的结果。/l/ 这张图靠前有很多点子，后头就没有。右边点子稍多一些，说明舌头左边比右边低，气是从左边出来的。/t/ 的舌头跟腭接触的地方比 /l/ 多，每一个电极都有接触。/s/ 和 /t/ 很象，但有一点基本不同：/t/ 的舌头完全堵住，/s/ 前头有个空缺，也就是舌头

上面有条沟,气就通过那里喷出来。

但是,这种电子腭图仍旧有一个很基本的缺点,就是一定要有接触，才能知道舌头的部位。有许多音，舌头并没有接触什么地方,只是因为舌头形状不同引起了声波不同。有什么办法可以把舌头的形状表现出来呢？最近我们的实验室在这方面的研究有一些新的进展。做这方面工作的是一位电气工程师,叫庄秋广。他是利用光的特点来做的,叫“光腭图”(optical palatagraphy)。也是做一个假腭,可是在假腭和舌头之间有一对一对的电极,每一对电极中的一个发出一条很细很细的光，另一个则受光。光有固定的速度。舌头离得比较近,光的反射就快；离得比较远，反射就比较慢。由光的反射速度,就可以测量出舌头和腭的距离。在我看来,这是一个相当有前途的仪器。

〔附〕 剪磁带的方法

做语音实验常常需要剪磁带。下面简单介绍一下剪磁带的方法。

录音的时候,比如录进一个 [p'a] 去，要想把它剪开，正好剪在 [a] 的开头和 [p'] 之间,这就相当麻烦。原来有两个办法。一个是用一种药水抹在磁带上，从药水的分布情况看出应该在什么地方剪开。但这常常看不清楚,而且容易损伤磁带。另一个办法是把放音磁头接在声谱仪或示波器上,慢慢移动磁带,看看什么时候有冲直条,什么时候有共振峰。但这要反复地做才能确定，因此也不很理想。

如果直接利用声谱图，就能够剪得比较准。可以先在 [p'a] 的前面或后面用磁铁性刀片在磁带上划一道痕，放音放到这道痕的时候,就可以听到“啪”的一声。然后再把这一段磁带做一个声谱图。在图谱上除了 [p'a] 以外,划的那道痕也有它的冲直条。从

图上量出从这个冲直条到要剪开的地方的距离，按时间刻度算出时间，然后根据磁带的速度计算，就可以量出该在磁带的哪个地方剪开。比方磁带的速度是每秒二十厘米，从声谱图上量出从冲直条到剪口是半秒钟的间隔，就在磁带上从刻痕开始量到十厘米，把它剪开，这样就可以剪得很准。（最好用非磁性的剪刀，例如铜的或塑料的，那样不至于干扰磁带。）

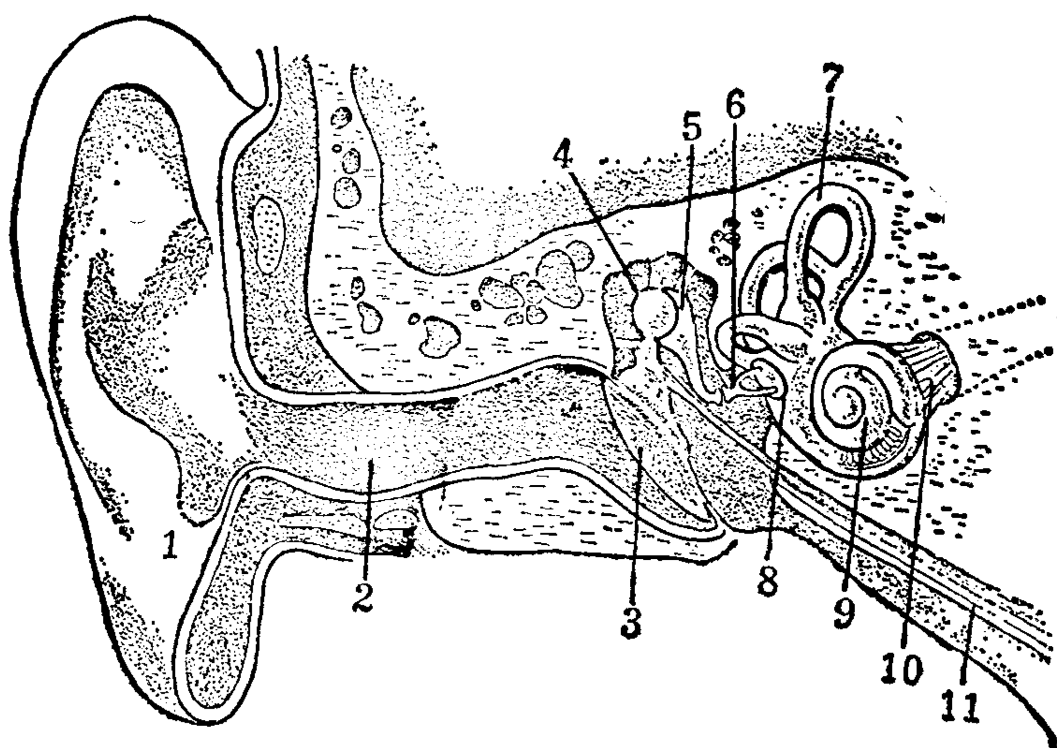
做这样的实验，如果磁带速度比较慢，剪下来的磁带就比较短，不好利用。最好把它的速度加快，磁带的长度就可以增加。如果再用更快速度的磁带转录，还可以把磁带的长度增加好几倍。

八 听觉

人耳的构造和功能

人的耳朵是一种非常灵敏的器官，能感觉到空气压力极微小的变化。它由外耳、中耳和内耳三个部分组成。研究语言的人应该对耳朵有一个基本的了解，因为没有耳朵就没有语音，没有语音就可能没有语言。下页图 8-1 是人耳构造图。

外耳看起来很简单，其实是很重要的。如果没有外耳，耳朵就绝对不会这样敏感，许多声音就会听不到。外耳不仅收集声波，而且还起共振作用。外耳道也是一根一头开着、一头关着的管子，和从口腔到声门的情况是一样的（它的尽头是鼓膜，所以是关着的）。既然是一根管子，就可以算出它的共振峰的频率。前面已经说过，共振峰的频率是声音的速度除以它的波长，波长则是由管子的长短决定的。外耳道长约 25 毫米，波长等于 4 乘以管子的长度。这样：$f=V/\lambda$，$\lambda=4\times 25\ \text{mm}=100\ \text{mm}$，$V=340$ 米（每秒），因此 $f=340{,}000/4\times 25=3{,}400\ \text{Hz}$，可见外耳道的共振峰是 3,400 赫左右。正因为外耳道的构造是这样，所以一般人对于



1. 耳廓 2. 外耳道 3. 鼓膜 4. 锤骨 5. 砧骨 6. 镫骨
7. 半规管 8. 前庭窗 9. 耳蜗 10. 听觉神经 11. 咽鼓管

图 8-1

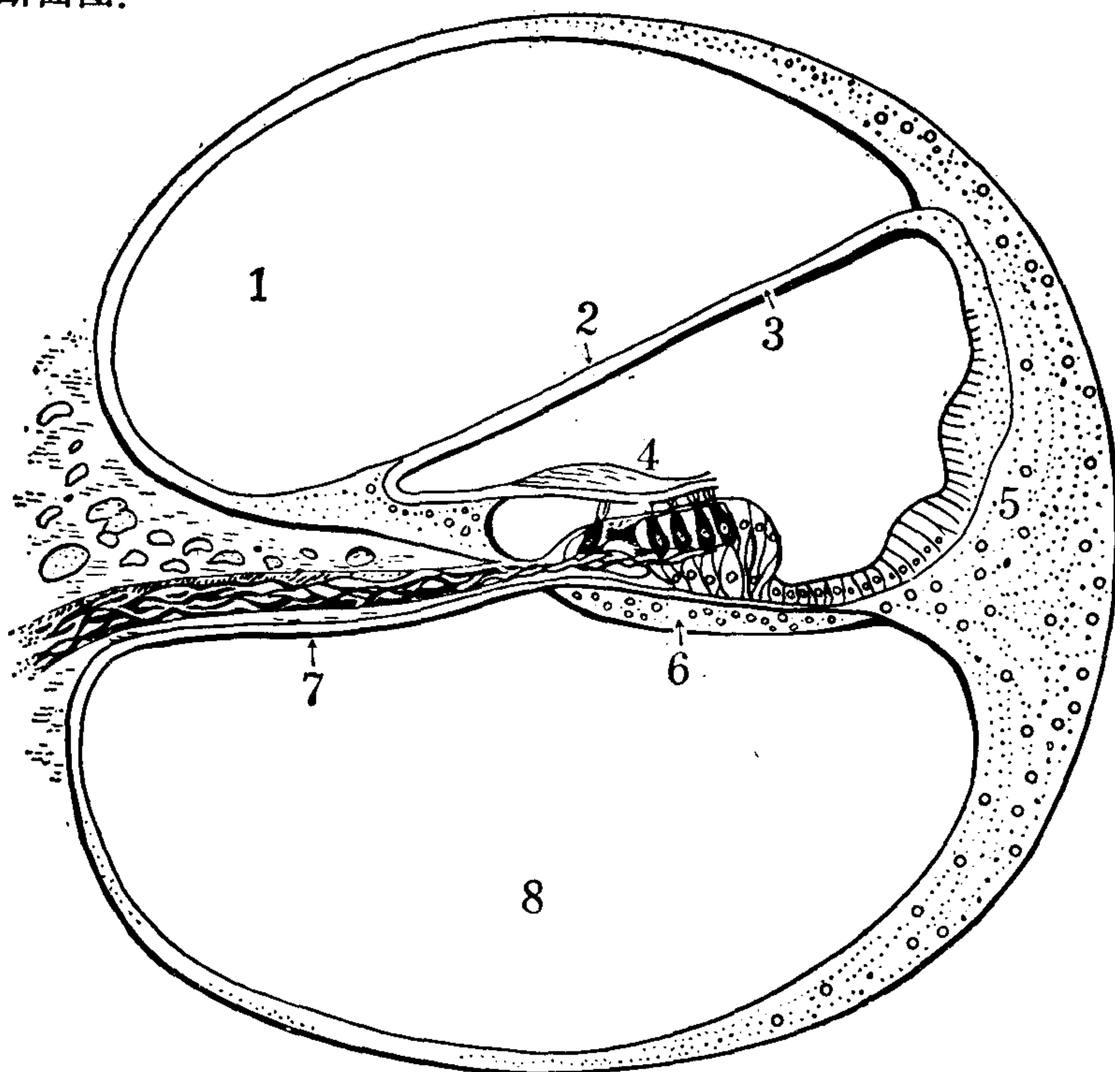
3,000 赫到 4,000 赫之间的声音最为敏感。

中耳是由鼓膜把它和外耳隔开的。鼓膜是一块椭圆形的稍向内陷的薄膜，面积约 0.65 平方厘米，处在外耳道的尽头。中耳里最主要的是三块比小指甲盖还小的听小骨，它们互相连接，根据形状依次叫做锤骨、砧骨和镫骨。中耳除了跟内耳相连以外，底下还有一根管子，叫咽鼓管，通到咽部，它可以调整气压，使鼓膜内外的压力保持平衡。

从语言的角度来说，听小骨的作用是扩大和增加空气的气压，让内耳受到更大的振动，使听觉更加灵敏。听小骨之所以能起这样的作用，是因为镫骨也是连接在一块膜皮上的。这块膜皮叫作前庭窗。它的面积比鼓膜小得多，因此，由鼓膜的振动而产生的气压，经过锤骨、砧骨、镫骨这三块软骨的传导到达前庭窗时，压力就增加了好多倍。而当声音太强、压力太大的时候，听小骨上的肌肉

又会把这三块软骨拉紧一些，减少振动的压力，以保护内耳。如果把电极插在听小骨的肌肉上，就可以发现这些肌肉随着声压的变化总在不断地活动，即使在人睡觉的时候，这些肌肉也是在动的。

内耳包括前庭窗、耳蜗和半规管。最主要的是耳蜗，它是转着卷起来的。蜗管里充满了很粘的淋巴液。图 8–2 是卷曲的耳蜗的横断面图：



1. 前庭阶　2. 前庭膜　3. 淋巴　4. 柯梣氏器官
5. 蜗管　6. 基底膜　7. 听觉神经　8. 鼓阶

图 8–2

蜗管内部分为两部分，由一层薄膜隔开，这层薄膜叫基底膜。基底膜上附着很多复杂的毛细胞，它们一组一组地构成了一个奥妙的

器官，叫作柯替氏器官 (organ of Corti)。在柯替氏器官上有很多很细的神经，这些神经出了蜗管就汇集成两条相当粗的听觉神经，通到大脑。大脑底下有十二对重要的神经，听觉神经是第八对。

空气的振动从外耳通过鼓膜经过听小骨传到前庭窗，前庭窗就随着发生压紧、松开的振动，使蜗管里的液体也发生振动，影响到基底膜。而基底膜受到影响那一部分的柯替氏器官上的神经就受到刺激，变成神经上的信号，传导到大脑。由于蜗管里液体所受到的压力是随着前庭窗的振动而发生变化的，所以基底膜上柯替氏器官所受的刺激也是有变化的。在哪一部分受到刺激，大脑就能够分析出那个声音的频率是多少。离前庭窗近，频率高；离前庭窗远，频率低。一个人分辨声调、元音和辅音，主要就是依靠频率。所以，频率这个概念从生理上说，就是基底膜柯替氏器官不同部位的振动在大脑里的反应。这就是频率这个概念的生理基础。

两耳听觉能力的差异

声波变成毛细胞的颤动，由听觉神经传达到大脑的时候，两耳传递的速度和刺激的强度是不同的。通常是右耳传到左大脑的速度比较快，刺激比较强；左耳传到右大脑的速度比较快，刺激比较强。换句话说，左大脑接受右耳声音比较灵敏，右大脑接受左耳声音比较灵敏。这种生理现象叫“对侧的”(contralateral)。和对侧的相反的就叫“同侧的”(ipsilateral)。人体的构造，尤其是神经系统的构造，大都是根据对侧的原则的。

六十年代以后，有不少人根据左右两耳和大脑的联系不同这一现象，对两耳在听辨语音上是否有所不同做过许多实验。这种实验叫做“两歧实验”(dichotic experiment)。例如，在同一时间内让左右两耳分别听到不同的数字，左耳听到 1 的时候，右耳听到的是 6，这样，两耳同时收到的信号是冲突的。在这种情况下，听

的人有时会把听到的数字说错，例如说成 9；更多的人则感到声音很乱，但是认为最可能是 6。一系列实验的结果表明，右耳一般要比左耳听得清楚。这是两歧实验得到的最基本的收获，叫做“右耳优势”(right ear advantage，简称 REA)。后来又有人用音乐来进行实验，发现情况和语音完全相反，右耳反而不如左耳，是“左耳优势”(left ear advantage，简称 LEA)。于是有很多人认为，这种现象说明大脑的两部分结构不同，分工不同：语音是由左大脑管的，所以右耳比较敏锐，是 REA；音乐是右大脑管的，所以左耳比较敏锐，是 LEA。

但是，情况并不这样简单。就以语音来说，元音和辅音的情况就不一样。两耳分开听元音，能力差不多；分开听辅音，右耳就比左耳强，是右耳优势。但除了元音和辅音，语音还有超音段成分。特别是声调，是一种音高，近似音乐，究竟应该是 REA，还是 LEA 呢？最近，加州大学洛杉矶分校有两位女研究生对这个问题用泰语做了两歧实验（比如，在同一时间内左耳听到 mā，右耳听到 mà，让左右耳的声调发生冲突）。她们研究的结果是声调和音乐不一样，是“右耳优势”。泰语的声调和汉语一样，是有区别词义的功能的，是语音系统的一个成分，所以和别的语音成分同样是 REA。它的强度和辅音差不多，比元音要强。

和上面介绍的实验相类似的还很多。从 1961 年开始到现在，大约已经有几百篇文章讨论过这方面的问题，这里不能详细介绍。下面介绍另一类很有意义的实验，是研究大脑和语言的关系的。

大脑和语言的关系

研究大脑和语言的关系的实验是最近才开始的。其中一种是脑血流实验。这种实验可以帮助我们弄清楚大脑哪些部分对语言是最重要的。

大脑的重量和人的整个体重相比是很小的，只占身体重量的五十分之一左右。可是，大脑对血液的需要量很大，心脏输送的血液，五分之一都是供大脑使用的。最近，通过实验，利用电子计算机进行计算，已经可以知道大脑的哪一部分在什么情况下用血最多。比如，让一个人坐着做不同的事情，比如说话，听音乐，听人讲话，等等，这时，研究人员测量他大脑里血流的变化情况，就会发现，人在做不同事情的时候，大脑里血流的波形、分量以及血流的过程都随着在起变化。反过来说，从血流图上就可以看出这个人在干什么，比如是在说话，还是在听音乐，等等。这方面的工作最近在瑞典和美国有三组人在做。瑞典的一个医院和一个语言学家合作研究了说话和听话时左大脑的血流情况。他们把左大脑分成许多小方块，用电子计算机把血流经过每个小方块的情况计算出来画成图。如果血流多，分量多，速度高，这部分区域在图上就会呈现出红色或白色；如果血流少，分量少，速度慢，就会呈现出蓝色或绿色。根据电子计算机画出来的血流图看，左大脑的卜洛卡区(Broca's area) 和维尔尼克区 (Wernicke's area) 对语言来说是最重要的。前一个区域在说话时起主要作用，后一个区域在听话、了解语义时起主要作用。以前必须动手术把头盖骨打开才能进行这方面的研究，现在由于知道了血流和脑子里行为的关系，用不着动外科手术就可以进行这方面的实验和研究了。

近几年来，有不少人用"脑电流图"研究语音和大脑的关系。用脑电流图实验的办法很简单。紧贴着头部装上几个甚至几十个电极，在使用语言(说话、听话等等)的时候，大脑里神经细胞之间发生交流，并且产生电流。通过电极，把电流测量出来，就可以画出脑电流图来。

这种实验最大的困难在于如何把大脑里语言的行为和别的行为分开，因为说话的时候同时也会看到和听到许多别的行为和声

音，这些活动在大脑里也是有反应的。为了从脑电流图里找到专门针对语言或语音的波形，最近采用了一种叫做“平均电压”(averaged evoked potentials，简称 AEP）的方法，那就是一次一次地重复某个发音。例如，只说一次 [ba]，是无法从脑电流图里把它和其他反应分开的。但如果 [ba] 的发音次数增加到三十二次，其他没有固定性的行为在图中就渐渐没有 [ba] 显著了。假如以三十二次作为一个平均单位取一个脑电流图，然后是六十四次，一百二十八次，五百一十二次，把这几张图上显示出的 [ba] 这一发音行为的波形加在一起，就可以研究出针对这一发音行为的脑波。

用脑电流图研究大脑和语言的关系是 1978 年才开始的。这种试验至少已经证明 E. Sapir 几十年前的一个猜想是对的。Sapir 在 1921 年出版的《语言》(Language) 一书里曾经提到：念 [p] 的时候，嘴唇的动作和吹蜡烛很相象，那么这两种肌肉上一样的动作，在大脑里的反应是不是也相同呢？Sapir 猜测那是很不一样的，因为吹蜡烛是一个孤立的动作，和语言毫无联系；而双唇塞音 [p] 是整个语音系统里的一个单位，和语义有密切的联系。最近有一位医学家和一位语言学家合作用脑电流图的方法做实验，证实了 Sapir 五六十年前的这个想法是正确的。他们让发音人头上戴上十几个不同的电极，有时候让他念许多 [p] 开头的词（如 plash, plate, pledge 等等），有时候又让他吹一张纸。从这两种行为的 AEP 对比中可以看到，吹纸的时候左大脑和右大脑的波形没有什么区别，但是在发 [p] 这个音的时候，左大脑的 AEP 就非常大，而右大脑就没有什么变化。两种嘴唇动作虽然很相象，但到底一个是语言，一个不是，在大脑里的反应也就不一样。

几种听觉实验

上面介绍的几种实验都说明大脑左右两部分的机能是不同

的。心理学家还曾经进行了一些实验，发现语音的右耳优势现象在三岁到六岁的儿童身上就有反映。因此有人认为，从学习语言的观点来看，五、六岁这一段时期是一个关键。婴儿刚出生的时候，神经系统之间有许多地方还没有连起来，这一条神经系统的信号还传递不到另一条神经系统上去。从一岁到三岁，神经系统发展很快，脑子从 30 多克长到将近 1,000 克，大部分神经系统之间的突触 (synapse) 也逐渐长成。到了七岁至九岁，这些突触就完全长好了，这时给他训练新的习惯，比四、五岁时要困难得多。此外，在解剖中还发现，刚生下来的婴儿，左脑有一部分（这部分的拉丁文名称是 planum temporale）就比右脑大。这也启示我们，人生来就有一些先天的物质基础为后来大脑两部分不同的发展做了准备。

下面介绍一种叫"范畴感知"(categorical perception) 的实验，它涉及人类听辨音位范畴的能力是否是先天的这个问题。

我们知道，元音是一种连续现象，不是分立现象。比如从 [i] 到 [e] 之间，就有无数过渡的元音，因为发音时舌头从 [i] 到 [e] 是连续的；要是测量共振峰 F_1 F_2 F_3……，从 [i] 到 [e] 也是连续的。所以，从物理学的观点来看，元音是无限之多的。可是世界上各种语言的元音总是有限的，也许是十几个，也许是几十个。我们说某个语言有多少个元音、多少个辅音的时候，实际上已经把无限之多的连续现象归纳成很少几个音位范畴了。这就是所谓"范畴感知"。范畴是音位学里最基本的概念。

人辨别信号的能力是有限的，不可能从无限多的语音中去识别各种不同的声音。关键是人能把无限多的语音归纳成有限的音位范畴。归纳范畴，就需要划范畴的界线。在不同语言的语音系统中，音位范畴的界线和数目都是不同的。比如在 [i] 和 [e] 之间，日语只要划一条界线就可以了，因为日语只有五个元音：[i]，

[e], [a], [o], [u]。可是英语就需要划两条界线，因为它还有一个[ɪ]。可见，在同一个语音区域里，日语只有两个范畴，英语就有三个范畴。近几年来，实验语音学关心的基本问题之一就是哪些语音现象是人先天就具备的，哪些是从后天环境里学到的。这类问题的讨论叫做“本能对教养”(nature vs culture)。本能是从遗传基因里得到的，是不需要学习的；教养则是从后天环境学到的东西。音位范畴的界线究竟是从本能得到的，还是从环境学到的呢？

在这方面最惊人的发现是1971年1月心理学家 P. D. Eimas 在美国《科学》(Science) 杂志上发表的一篇文章。文章的题目是《婴儿的言语感知》(Speech Perception in Infants)。Eimas 研究的是 VOT (voice onset time, 可译成清浊对比度，见第56页) 里的音位界线问题。他的实验室的工作人员测量和统计了好几十种语言的 VOT, 发现尽管各种语言里 VOT 的情形各不相同（例如俄语 [p] 的 VOT 起点很早，英语就比较迟，汉语更迟），但是，VOT 是能起区别音位的作用的，它能起区别作用的信息总是出现在一定范围之内的。比如，假定 VOT 是 −60，在 −60至 −40 毫秒这一段时间内，从来不出现可以区别音位的成分；可是从 −40 到 −20 毫秒，同样也相差20毫秒，在有些语言里这一段时间就很重要，因为可以算区别性成分了。换句话说，在 −60 到 − 40毫秒这段时间不可能有音位界线，而在 −40 到 −20 这个同等分量的物理量之间就可以有音位界线。这种区别音位界线的机能是否先天就有呢？

为了探索这个问题，Eimas 把婴儿作为研究对象。刚出生几个星期的婴儿是一组，出生三、四个月的婴儿是另一组，分别进行实验。他们注意到，婴儿在吸吮奶头的时候，从开始起经过一段时间，速度就会固定下来。于是就先测量出每个婴儿用橡皮奶头吸桔子水的速度（每秒钟吸多少次），然后突然开灯、开收音机或让他

们看一个玩具，给他们一个新的刺激。这时婴儿吸桔子水的速度就会增加。但是过了一段时间，对新的刺激习以为常了，吸桔子水的速度就又恢复正常了。根据婴儿的这种表现，Eimas 用言语合成的办法合成了几个音节，其中的 VOT 各不相同。我们可以用 S_1, S_2, S_3 来代表这几个音节。$S_1 \leftrightarrow S_2 \leftrightarrow S_3$ 依次排列，相隔的时间是相等的，从 S_1 到 S_2 中间是没有音位界线的，从 S_2 到 S_3 中间是有音位界线的。婴儿吸桔子水的时候，就在旁边替换着播放这三个音节。实验的结果发现，当从 S_2 换到 S_1 的时候，婴儿吸桔子水的频率并没有变化；当从 S_2 换到 S_3 的时候，吸桔子水的频率就会提高，过了一段时间才慢慢降下来。这说明，虽然从物理观点看，$S_2 \leftrightarrow S_1$当中的区别，和 $S_2 \leftrightarrow S_3$ 当中的区别是一样的，但是 $S_2 \leftrightarrow S_1$ 之间没有范畴的界线，也就是说它们是同一个范畴，因此不会对婴儿产生新的刺激，而 $S_2 \leftrightarrow S_3$ 之间有范畴的界线，所以从 S_2 转到 S_3，对婴儿的听觉就是一个新的刺激。这就叫做“范畴感知”。这个实验相当精彩，在以前很难想象这么小的婴儿就可以用来做这类实验。他们的实验报告发表以后，又有一些语言学家和医院合作做了一些类似的试验，比如测量婴儿心跳的速度或出汗的速度，等等，方法虽然不同，但是也都得出了同样的结果。

下面介绍另一种叫做“选择适应”(selective adaptation) 的实验，这种实验和范畴感知的实验是密切相关的。实验的内容是转接和音位界线的关系。这种实验分为两部分。第一部分叫辨认 (identification)，第二部分叫区分 (discrimination)。

第七讲已经谈到，转接是听辨辅音的主要信息，不同的辅音对元音的共振峰有不同的影响。以 [ba] 和 [da] 为例，元音都是 [a]，但是由于 [b] 和 [d] 不同，所以 [a] 的 F_2 开头的指向也不同，[b] 的指向低，[d] 的指向高。我们可以画出许多不同的转接，如下面图 8-3。做第一部分辨认实验的时候，先把这些不同的转接

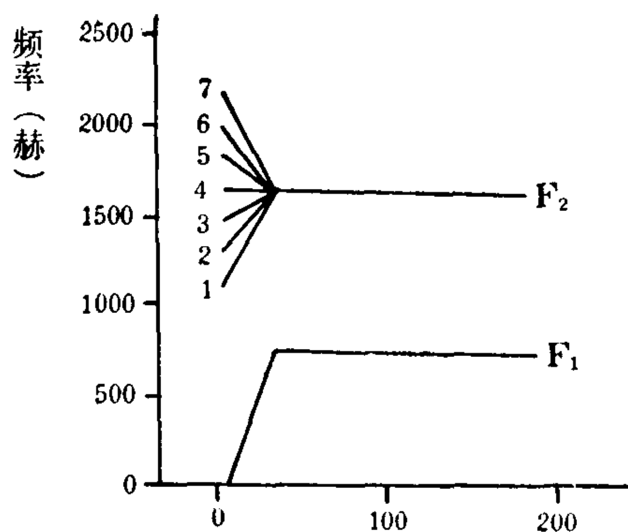


图 8-3

编上号码，如图 8-3 里的 1,2,3,4,5,6,7,然后让一些人听辨每一个音节是 [ba] 还是 [da]。这样就可以发现,转接最低的时候,回答是 [ba]的比率是百分之百。转接逐渐升高，回答是[ba]的比率就逐渐下降,回答是 [da] 的比率自然就逐渐升高。转接升高到了一定的界线,就变成百分之百是 [da] 了。如果再继续升高,回答是 [da] 的比率又逐渐下降,也就是说, [da] 听起来越来越不清楚了。以上这种变化可以用图 8-4 来表示。图中横轴代表转接的变化，纵轴代表辨认的百分比。带▲的线代表 [b]，带●的线代表 [d]。两线相交的地方,听的人最没有把握,觉得有时候象 [ba],有

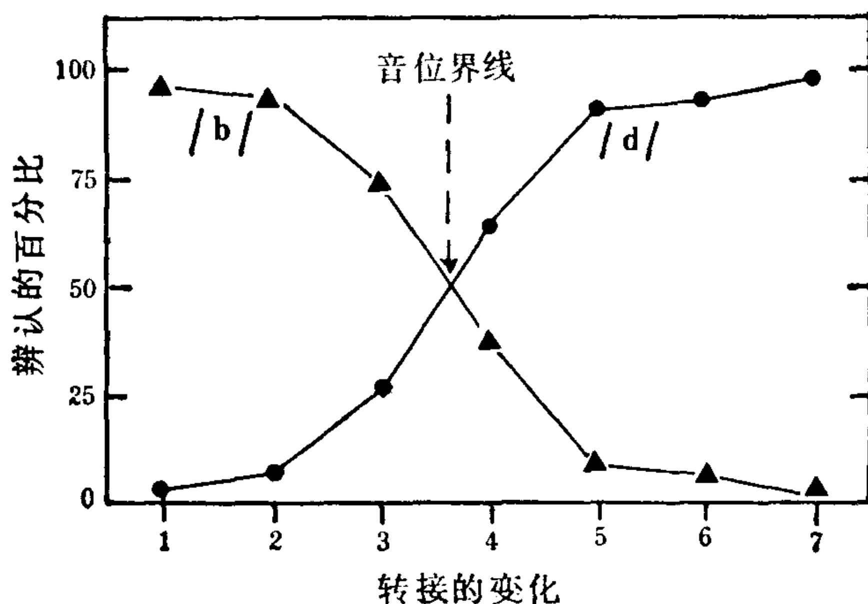


图 8-4

时候象 [da],因为它既不在 [ba] 的范畴里,也不在 [da] 的范畴里。

从实验的角度看，这个交界的地方可以算是 [b] 和 [d] 两个音位的界线。

第二部分区分实验有几种不同的做法。有一种程序叫 ABX。每一次让人听三个音节，第一个音节A和第二个音节B总是不同的，第三个音节一定是前两个音节中的一个(A 或 B)，听的人的任务就是回答第三个音节 X 到底是 A 还是 B，所以这种方法就叫做 ABX。实验的结果表明，回答的准确率是百分之五十的时候，到了音位界线它就向上升，过了音位界线就又下降了。可见音位界线不但在辨认实验里可以找到，就是在区分实验里也是可以找到的。我们的听觉在音位界线那个地方最敏锐、最清楚。在某个范畴之内听觉可能不怎么敏锐，但是在两个范畴之间的界线上，我们就听得特别准。

从以上这些实验可以证明，音位界线是先天就有的。我们还可以进一步追问，是不是所有的音位界线在听觉中都是先天就有的？还是有的是先天就有的，有的是后来从环境里学到的？世界上有几千种语言，语音系统各不相同，所用的区别性特征也各不相同，哪些是和先天有关的，哪些是没有关系的？针对这类问题，我们在加州大学作了一些实验。我把实验的结果写在《语言变化》(Language Change)这篇文章里，发表在 1976 年《纽约科学院年刊》(Annals of The New York Academy of Sciences)上。

这个实验是以普通话的声调——阴平和阳平作为材料的。这两个声调之间也有连续的现象，就象前面实验里的 VOT 或 [b] 和[d] 的转接一样。我们先用语音合成的办法做出十一个音节，用的是元音 [i]，从阳平开始，每个音节的音高斜度逐渐减小，越来越平，一直到 55 调阴平为止。因为是用仪器合成的，所以十分准确。每个音节的发音时间是 500 毫秒，开头五分之一的 100 毫秒是平的。第一个音节的 F_0 一开始是 105 赫，经过 100 毫秒以后就往

上升，到 500 毫秒的时候就升到 135 赫，也就是说，在 400 毫秒时间内上升了 30 赫。第二个音节的 F_0 比第一个音节高 3 赫，两者的区别是△F，从第二个音节开始，依次加上去，就是 108 赫，111 赫，114 赫，……一直加到第十一个音节 135 赫，如图 8-5。

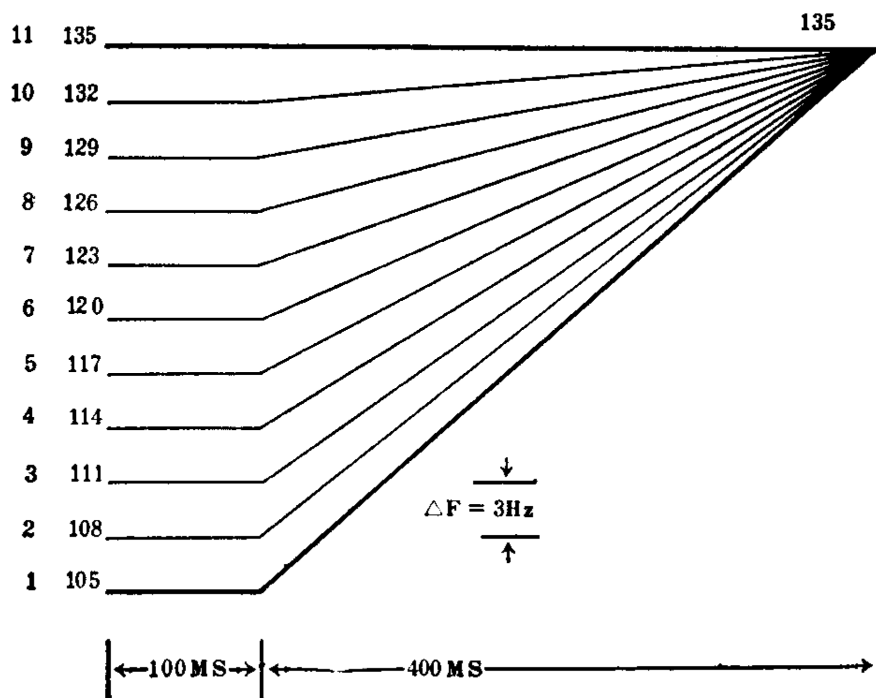


图 8-5

前面已经说过，做这类实验分两个阶段，第一阶段是辨认，第二阶段是区分。我们所设计的实验，在辨认阶段，每个音节是半秒钟，然后有 4.5 秒时间让听的人回答和休息。在区分阶段，用的也是 ABX 方法，A 和 B 之间相隔 0.4 秒，B 和 X 之间相隔 0.6 秒，休息一下回答，加在一起一共差不多是 4 秒。同我们合作参加测验的人一共分两组。第一组是几个中国人，第二组是一个美国人。这个美国人完全不懂汉语，可是有很好的音乐训练，对音高平不平听得很准。在辨认实验阶段，我们就要求他回答是平还是不平。实验结果，他的平和不平的界线是在第十和第十一音节之间，也就是说，只有那个最高的第十一个音节（135 赫）他听起来才是

平的，其他都认为不平。他的听觉和物理现象是完全一致的。用ABX 的方法做区分实验的阶段也是一样。一般情况他说对的百分比总是在百分之五十左右，但是到了音位界线，他的区分能力一下子升得很高。图 8-6 是实验结果，带 ∘ 的线代表辨认实验，带 + 的线代表区分实验。

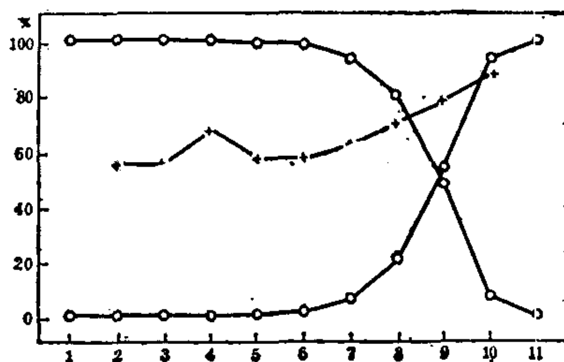


图 8-6

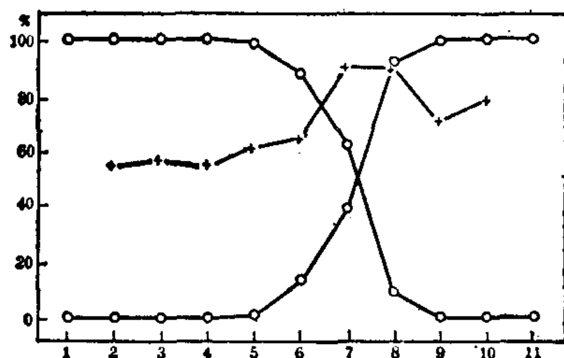


图 8-7

另一组中国人的情况就不一样，他们的音位界线都在第七个音节左右。换句话说，这个音位界线不可能是先天就有的，这个阴平和阳平的界线是从环境里学到的。请看左图 8-7。

下面图 8-8 也是一个中国人的测验结果。他是研究汉语的心理学家，在这方面有多年的工作经验。他的图是很有特点的。和图 8-7 比较，辨认实验的结果是差不多的，音位界线都在第七个音节左右。但是区分实验的结果很不一样。由于他有多年这方面工作的经验，耳朵比一般中国人灵敏，因此他的区分能力在第七个音节，也就是中国人的音位界线

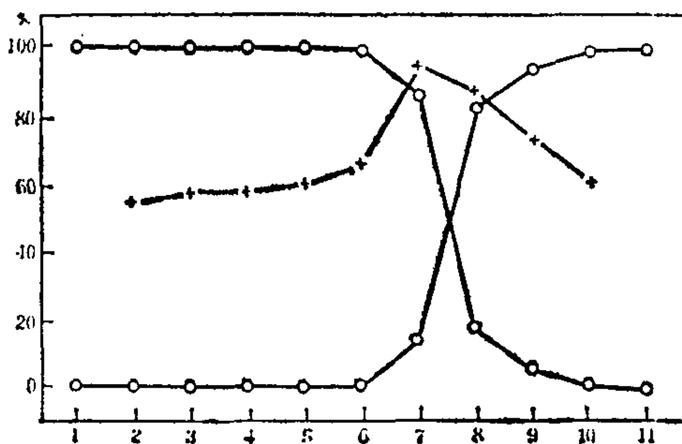


图 8-8

上。这显得特别突出。他对音高的平不平也听得比一般人准确。

由以上这些实验可以看出，从听觉的角度看，音位范畴内部和音位范畴之间的差别是很大的。可以用图 8-9 来说明。两个圆圈代表两个音位范畴。在范畴内部的变化（A'↔B'）是很难区别的，而 A↔B 虽然和 A'↔B' 的距离一样，但因为 A 和 B 这两点出于两个不同的范畴，因此 A↔B 就很容易区别。

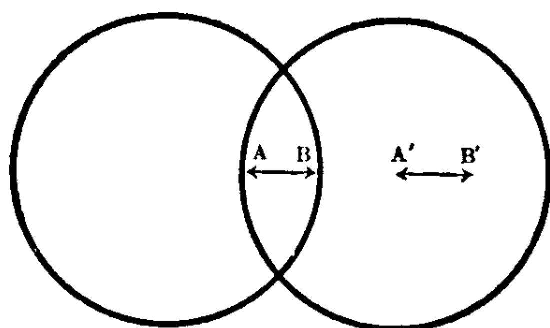


图 8-9

根据以上介绍的一些实验，我们可以认为，有些音位界线的确有它先天的生理基础。从几个星期到几个月的婴儿，虽然还没有学会任何语言，但是从他们的大脑情况来看，已经可以区别什么是范畴之内的，什么是范畴之外的了。同时，我们也可以认为，有些音位界线并不是先天就有的。比如普通话的阴平和阳平，实验证明，它们之间的音位界线是人们从环境里学到的。

最后再介绍一个实验。它是接着范畴感知的实验做的。比如说，用范畴分界（categorical boundary）的办法，把 [b] 和 [p] 的界线找出来。让听的人戴上耳机，在一秒钟内高速度连续放送好几个 [p]，加强对他听觉的刺激。然后拿掉耳机，再做范畴分界的实验。这时可以发现，[b]和[p]之间原有的音位界线往[b]的方向移动了。[p] 的刺激越强，界线就移动得越多。这是因为每一个区别性特征在大脑里都有专门听它的神经。[b] 和 [p] 的区别性特征的不同就在于清浊，也就是 VOT 不同。在强刺激的情况下，听 [p] 的那一部分神经过于疲劳，因而不那么敏感了，所以它和 [b] 之间的音位界线就移动了。有人认为这是区别性特征在生物学上的证明。音位学已经有好几十年历史了，但是它和物理学、心理学等

等联系起来，还不过是近几年的事。如果我们能够给每一个区别性特征不但找到社会上的证明，而且还能够找到人体生理上心理上的证明，那将是一个非常巨大的贡献。

〔附录一〕 关于区别性特征理论

一般语言学家往往把语音研究分成两个方向：一个是语音学(phonetics)，另一个是音位学或音韵学(phonology)。在音位学里，区别性特征理论(distinctive feature theory)是一个很基本、很常用的理论。

区别性特征理论可以说有两个不同来源。一个是捷克的布拉格学派(Prague school)。他们比较重视语音学和音位学。那个时代最有影响的是N. Trubetzkoy在1939年写的《音位学原理》(Grundrüge der Phonologie)，这本书直到最近几年才译成英文。Trubetzkoy和R. Jakobson是好朋友，从他的书里可以很清楚看到Jakobson的影响。另一个来源是美国的E. Sapir。Sapir在1925年写了一篇《语言的音型》(Sound Patterns in Language)，这是非常有启发性的一篇文章。Sapir认为每一种语言都有自己的音型。音型分成内部格式(inner configuration)和外部格式(outer configuration)两部分。内部格式是比较抽象的，是分析一种语言的时候给它构拟出来的；外部格式就是大家都可以听得到、量得到的东西。分析语音就是想办法从它的具体表现构拟出它的抽象系统，这个系统就是语言的音型。

近些年来，最重要、最有启发性的一本书是1951年R. Jakobson, M. Fant和M. Halle合写的《语音分析初探》(Preliminaries to Speech Analysis)。Halle是Jakobson到美国以后的学生，他一方面受到Jakobson从布拉格学派带来的影响，另一方面又受到美国本土Sapir的影响。他的博士论文是Jakobson指导

写成的，题目是《俄语的音型》(The Sound Pattern of Russian)。1968 年他又和 N. Chomshy 合写《英语的音型》(The Sound Pattern of English)。这两本书都非常重要。两本书都用"音型"作书名，说明他们对 Sapir 的尊敬。Halle 和 Chomsky 合作以后，希望把音位学纳入当时的生成语法理论里去，因此这种理论就被称为"生成音位学"(generative phonology)。在过去二十多年里，国外的音位学，特别是美国的音位学，最重要的发展就是生成音位学。在这方面 Halle 的贡献很大，但是他不象 Chomsky 那样爱活动，只是做学问，因此不象 Chomsky 出名。

区别性特征理论和传统语音学有很多基本不同的地方。第一，传统语音学往往把一个音段(如æ, n)看成是语音的最小单位，不再继续分析下去。区别性特征理论则认为，一个音段固然在时间上是最小单位，但作为语言结构里的一个成分，它并不是最小单位。音段是由好几个区别性特征组合起来的，区别性特征才是最小的单位。他们对区别性特征做了系统的叙述，比如，一共有多少个，它们之间的关系如何，等等。第二，传统语音学差不多都是把元音和辅音分开谈的，好象元音和辅音没有太大的关系。区别性特征理论认为这是不妥当的，因为不管发什么音，总是舌头和嘴在动，总是同一套发音器官，所以元音和辅音应当用同一套区别性特征来表达。所有不同的音段，都可以画出一个区别性特征矩阵(distinctive feature matrix)。

和传统语音学比较，区别性特征理论有许多进步的地方。它可以很简单、很容易地表达出：1. 自然类 (natural class); 2. 中立化 (neutralization); 3. 标记成分 (markedness)。下面分别做一些简单的介绍。

1. 自然类：根据不同的音的不同关系，可以分成若干个自然类。比如[i, u, m, p, k, s]这六个音段，[i, u]是一个自然类，因为

都是元音;[u, m, p]是一个自然类，因为都是唇音;[m, p, k, s]是一个自然类，因为都是辅音;[p, k, s]是一个自然类,因为都是清辅音。也有不成一个自然类的，比如[i]和[m]。这六个音段一共可以组成多少个类，用统计方法可以很快地算出来。大约总有好几百个。其中有的是自然类，有的不是。根据什么来确定呢？凡是有相同的区别性特征就是一个自然类，否则就不是。而这如果单用传统的国际音标,就很难看出来。其实,这种看法在几十年前就已经有了。二十年代,音位学布拉格中心 (Prague Circle of Phonology) 的一些捷克语言学家，象 N. Trubetzkoy, R. Jakobson, V. Mathesius, J. Vachek 等等，都很活跃，他们已经讨论到自然类的问题了,只是Jakobson, Fant 和 Halle 后来把它系统化,成了区别性特征理论的一部分。

2．中立化:我们可以举具体的例子来说明。美国英语有十二个单元音,三个复元音,元音系统相当丰富。但是,并不是在任何环境下整套元音都有区别功能。在 b—t 的环境里，许多元音都有区别功能（如 beat [bit], bit [bɪt], bet [bet], boot [but], but [bʌt], bought [bɔt], bat [bæt], bate [beit] boat [bout] 等等)。但在 -r 的前面,许多元音的区别就被中立化了,通常是一对一对地中立化了。比如[i]和[ɪ],在许多词里都有区别功能（如 peat 和 pit, beat 和 bit, lead 和 led 等等)，但是在 -r 的前面就没有区别功能。元音的区别功能在 -r 前面差不多减少了一半。再举两个例子：普通话的四声在很多情况下都有区别功能，但有时也有中立化的现象。比如在上声前，阳平和上声的区别功能就被中立化了。在上声前,只有三个不同的声调,不可能有四个,因为上声前面的上声是要变阳平的。英语 spin 一共四个音段,如果把每一个音段的区别性特征找出来，填在矩阵里，假定一共有三十个区别性特征，这个矩阵就是长四条，宽三十条，一共有 120 个

“+”“-”符号。但是英语的辅音并不是任意连在一起的，它有比较严格的规律。比如除了很少的词以外，如果一个词开头有两个阻塞音，前面一个一定是 s。(德语里有 ʃp，有的语言有 ft，有的语言有 hs，这在英语里都不可能。) 既然知道了这个情况，在矩阵里 s 这一整行的“+”“-”号就是多余的了。唯一需要知道的是这个音是阻塞音，这样所有其他的区别性特征都可以预先知道，它们的中立化就很自然地可以表达出来。

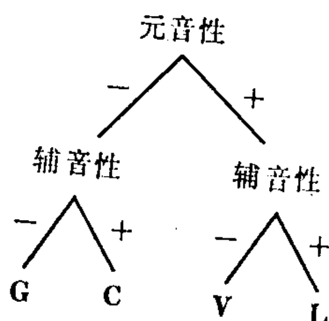
3. 标记成分：我们用高元音做例子来说明。高元音一般有 [i, u, y, ɨ, ɯ, ʉ] 六个。但在很多语言里只有两个。如果是两个，百分之八十可以肯定是 [i] 和 [u]，俄语、日语、西班牙语、意大利语都是这样。如果是三个，那就是 [i,u,y]，而不可能是别的，汉语普通话、德语、法语都是这样。这就有了两个问题：一是怎样把这种观察系统地表达出来，二是为什么会有这样的现象。区别性特征理论是用标记成分的方法把这类现象观察出来的。在这种矩阵里，并不是只用“+”和“-”，还要进一步用“m”表示有标记的(marked)，用“u”表示无标记的(unmarked)。这个理论的意思就是：每个语言都是在它的语音系统里想办法减低标记成分的。语言变迁的时候，总是从有标记成分变成无标记成分。小孩学话，总是先学无标记成分的语音，然后才一步一步学会有标记成分。患失语症的人一般是最先丧失有标记成分，无标记成分保存得最好。

按照区别性特征理论研究语音，就一定要先分出区别性特征。到底有多少区别性特征，现在还没有一致的意见。Jakobson, Fant 和 Halle 在 1951 年提出一共有十二个，这显然是不够的。后来陆陆续续增加，到 Chomsky 和 Halle 的《英语的音型》里，有三十多个了。但直到现在，还没有一套完全被大家接受的区别性特征。我们现在选几个比较常用的区别性特征，把一些音按不同的特征填成矩阵：

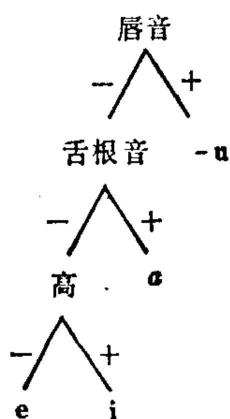
	i	ɑ	u	l	m	n	b	p	k	s	z
元音性 (vocalic)	+	+	+	+	−	−	−	−	−	−	−
辅音性 (consonantal)	−	−	−	+	+	+	+	+	+	+	+
浊　音 (voiced)	+	+	+	+	+	+	+	−	−	−	+
鼻　音 (nasal)	−	−	−	−	+	+	−	−	−	−	−
塞　音 (stop)	−	−	−	+	+	+	+	+	+	−	−
唇　音 (labial)	−	−	+	−	+	−	+	+	−	−	−
舌根音 (velar)	−	+	+	−	−	−	−	−	+	−	−

这样填完以后，每一对音段都应该能够区别出来，也就是说，任何一对音段至少要有一个区别性特征的符号是相对的，这就是“可区别条件” (distinctness condition)。列完了可区别条件，就形成区别性特征矩阵。

有了矩阵，就可以进一步画出区别性特征树形图了。一开始可以把所有的音段分成四大类：流音(L)，元音(V)，一般辅音(C)，过渡音(G)。这四类可以画成左边的树形图。



每个矩阵里都有许多多余的符号。把矩阵画成树形图，最主要的作用就是可以把多余的符号省去。比如 [i, e, a, u] 这四个元音，在上面的矩阵里有许多符号，画成树形图就成为左下图的样子。用这样的方法也可以把辅音区别开。如果从树根开始一直到最后，每次只有一个音段，那么可区别条件也就达到了。在这种情况下，自然对 (natural pair) 就可以比较容易表现出来。



一个自然对有几个不同成员。每一次指定一个自然对里的一个成员的时候，如果所需要的区别性特征比指定整个的对还多的话，那就是自然对。比

如 [m] 和 [n] 就是一个自然对：

	m	n
con.（辅音性）	+	+
nas.（鼻音）	+	+
voc.（元音性）	+	+
⋮	⋮	⋮
lab.（唇音）	+	−

要是指定这两个音段的话，所需要的是除了唇音 (lab.)以外的那些区别性特征。说简单些，如果区别性特征一共只有三个，那就是：

$$\begin{bmatrix} +\text{con.} \\ +\text{nas.} \\ +\text{voc.} \end{bmatrix} = \{\text{m,n}\}$$

如果要指定的是这个自然对里的一个，比如指定[m]，那么，所需要的区别性特征就是：

$$\begin{bmatrix} +\text{con.} \\ +\text{nas.} \\ +\text{voc.} \\ +\text{lab.} \end{bmatrix} = \{\text{m}\}$$

比较一下就可以发现，指定 [m] 需要四个区别性特征，而指定整个的对只需要三个，所以 {m,n} 就是一个自然对。相反，{m,i}就不是自然对。因为要指定{m,i}，需要把 [m] 的许多区别性特征和 [i] 的许多区别性特征都包括进去，区别性特征要增加许多。如果只指定其中之一([m] 或是 [i])，所需要的区别性特征反而比较少，因此{m, i}并不是自然对。

小孩学话往往也能反映出自然对的关系。比如，在学会发鼻音的时候，往往是好几个鼻音在很短时间内一起学会。音变也往往和自然对有关。比如元音鼻化，或者都是高元音，或者都是低元

音，不会是一个高，一个低，一个前，一个后。所以，自然对这个概念无论从哪方面来看，在语音上都是很基本的，很重要的。使用这个概念不能说空话，一定要具体，客观。这样即使用错了也容易被人看出来，因而也容易改正。

生成语音理论有一部分是讲特征、矩阵、树形图的，另外还有一部分是讲怎样写音位规则 (phonological rules) 的。下面介绍写音位规则的基本内容。先用一个普通话的例子来说明。

普通话最复杂的韵母是由前后两个过渡音 (glide) 和中间一个元音构成的，形成了 G_1VG_2 音丛。例如：

G_1	V	G_2	
i	a	u	“要”“跳”“票”
u	a	i	“歪”“乖”“摔”

但是，iai 这样的音丛就很少见，*uau 更是没有。由此可以看出，G_1 和 G_2 是起了异化作用的，两者总是相反的。这里包含一条规则，叫“剩余规则” (redundancy rule)。比如上面所说的异化现象可以写成：

$$[G] \rightarrow [+pal.] \ / \ [-pal.]\ V+$$

第一项指要变的音段（这里是 [G]），箭头代表变化，箭头后第二项指要变出来的区别性特征（这里是 [+pal.]），斜线后第三项指变化的条件（这里是 [−pal.] V+）。

前面已经谈到，一个词的各音段可以用矩阵表示出来。比如“乖” [ɡuai]：

	ɡ	u	a	i
元音性 (voc.)	−	+	+	+
高 (high)	+	+	−	+
腭音 (palat.)	−	−	−	+

剩余规则就是说，如果一个元音（V）的前面是一个非腭音（－palatal）的过渡音，那么，如果元音后面（＋）也是过渡音的话，它的符号就是多余的。因为它的条件已经被规定了，前头的既然是“－”，它就必然是“＋”。我们还可以写成和上面相反的情况：

[G]→[－pal.]/[＋pal.] V＋

如果元音前面是一个腭音（＋palatal），那么后面的过渡音必然是非腭音（－palatal）。

有了这两条规则，iai 和 uau 就不可能存在了。我们还可以看出，这两条规则是有固定关系的，所以还可以再简化一步，把它们合并。简化的时候一般用 α，β，γ 等希腊字母，比如我们现在用 α：同化的时候用 α/α，异化的时候用 α/－α。因为一共只有“＋”“－”两个符号，所以两条规则很容易合并：

[G]→[－α pal.]/[α pal.] V＋

如果后头的 α 是“＋”的话，前头的 α 也是“＋”，[－＋pal.]＝[－pal.]；如果后头的 α 是“－”的话，前头的 α 也是“－”，[－－pal.]＝[＋pal.]。这样就把这两条规则合并了。

印欧语的词头和词尾比汉语多得多。这些词头和词尾往往产生同化作用或异化作用，都是可以用音位规则来说明的。

比如英语的动词过去式 -ed，在清音后面是 [t]，例如 walked；在浊音后面是 [d]，例如 hugged。名词的多数在清音后面是 [s]，例如 cats；在浊音的后面是 [z]，例如 dogs。这种词尾读音的变化在英语里很常见，都是很简单的同化作用，可以写成这样的规则：

〔词尾〕→[α voiced]/[α voiced]＋

如果词尾前音段的 α 是“＋”，词尾变化的 α 就也是“＋”，也就是说，都是浊音（＋voiced）。如果前面的 α 是“－”，词尾的 α 就也是“－”，也就是说，都是清音（－voiced）。

再举一个词头的例子。英语的词头 in- 由于后接辅音的影响,有三种不同读音: in-受 p 影响读成 [im] (实际也写成 im, 如 impossible); in+telligent 的 in- 受到 t 影响仍旧读 [in]; in+congruous 的 in- 受 c [k] 的影响读成 [iŋ]。这也是简单的同化作用,可以写成:

$$[+\text{nas.}] \rightarrow \begin{bmatrix} \alpha & \text{pal.} \\ \beta & \text{lab.} \\ \gamma & \text{vel.} \end{bmatrix} \Big/ + \begin{bmatrix} \alpha & \text{pal.} \\ \beta & \text{lab.} \\ \gamma & \text{vel.} \end{bmatrix}$$

词头鼻音 (nasal) 的读音是由后面词根开头的辅音决定,或者都是腭音 (palatel), 或者都是唇音 (labial), 或者都是舌根音 (velar), 不可能不一致。

再举一个比较复杂的例子。在很多语言的 C_1VC_2 这样的结构里, 音节开头 C_1 的辅音系统比音节末尾 C_2 的辅音系统总是复杂得多。同一个辅音音位, 处在 C_1 和处在 C_2 也往往读音不一样。比如丹麦语,音节前的 [t] 到了音节后就变成 [d], 音节前的 [d] 到了音节后就变成 [ð]。这显然是同一种现象,可以列成下面的矩阵:

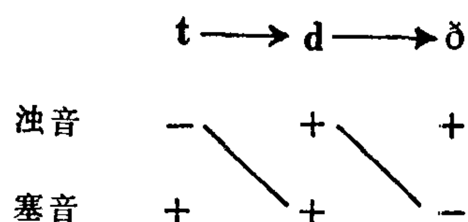
	t——→	d——→	ð
浊音 (voiced)	−	+	+
塞音 (stop)	+	+	−

这种变化比较复杂, 但是也同样可以写成一种规则。通常写规则总是有三个部分。就是:

〔输入〕→〔输出〕/〔条件〕

从前面列的区别性特征矩阵来看, 浊音一行的输出部分都是"+"号,可以很容易确定。塞音一行的输出部分比较复杂一些,要由输

入部分的另一个区别性特征浊音来确定：如果输入的浊音是"－"，输出的塞音就是"＋"；如果浊音是"＋"，塞音就是"－"。这种关系可以在矩阵里表现出来。（见右图）这是一种异化作用，比前面的例子复杂，因为它们的变数 (variable) 是画在不同的区别性特征上的。可以写成下面的规则：



$$[c] \rightarrow \begin{bmatrix} -\alpha & \text{stop} \\ \alpha & \text{voiced} \end{bmatrix} \Big/ \quad [\alpha \quad \text{voiced}]$$

在这条规则里有同样作用的两个音变。比如，d→ð 在这条规则里是怎样表现的呢？d 既然是浊音，α 是"＋"，变出来的就必然也是浊音。但是，不会是塞音，因为规则里塞音之前是－α。变出来的音既然塞音是"－"，浊音是"＋"，根据矩阵，可以确定它是 ð。

在规则当中还有一个相当重要的现象，那就是，规则和规则之间往往是有一定次序的。比如有两条规则 A 和 B，如果先用 A 后用 B 得出来的结果符合语音现象，而先用 B 后用 A 得出来的结果不符合，就说明 A 和 B 这两条规则是有次序的。这种现象 Sapir 和 Bloomfield 在很早以前就发现了，不过 Halle 把它系统化了。我们现在也举一个简单的例子来说明。下面两个词在美国英语里元音长短是不同的：

write $[ra^{i}t]$　　ride $[ra^{i}\text{:}d]$

以前我们说过，辅音对元音有固定的影响，比如，在 C_1VC_2 里，C_1 较多影响音高，C_2 较多影响音长，清辅音 ($\underset{\circ}{C}$) 和浊辅音 ($\underset{\vee}{C}$) 对元音长短的影响也不同。象上面的例子里 ride 的元音变长的现象，可以写成下面的规则：

$$V \rightarrow V:/+\underset{\vee}{C} \qquad (1)$$

这是第一条规则。如果在这两个词后头加上词尾 -er，成为 writer 和 rider，读音就有了新的变化，因为有这样一条语音规则：

$$\left\{\begin{matrix} t \\ d \end{matrix}\right\} \rightarrow \text{ɾ}/V+V \qquad (2)$$

例如美国英语 atom 就读成 [aɾəm]，而不读成 [atəm]，就是因为 [t] 处在 [a—ə] 之间。如果只根据这条规则：writer [ra^{i}tər]→[ra^{i}ɾər]，rider [ra^{i}dər]→[ra^{i}ɾər]，读音就变得完全一样了。但是很多美国方言，不但能区别 write 和 ride，而且也能区别 writer 和 rider，writer 读 [ra^{i}ɾər]，rider 读 [ra^{i}:ɾər]。也就是说，元音受 C̬ 影响变长的规则还是保留的，是先起了作用的。如果先用了 $\left\{\begin{matrix} t \\ d \end{matrix}\right\} \rightarrow \text{ɾ}$ 的规则，前一条规则 V→V: 的条件就不存在了。所以，必须是先用(1)，后用(2)：

		根据(1)	根据(2)
writer	[rá^{i}tər]⟶	[rá^{i}tər]⟶	[rá^{i}ɾər]
rider	[rá^{i}dər]⟶	[rá^{i}:dər]⟶	[rá^{i}:ɾər]

近些年来，对语音的叙述就是象上面所介绍的那样，相当抽象，也相当有系统。一般讲来，要叙述一种语言的语音系统，总是要分好几个步骤。第一步必须弄清楚这个语言一共有多少区别性特征。各语言里具有区别功能的特征是不相同的。然后就要做一个区别性特征矩阵。矩阵的系统可能不够严密，于是画出树形图。下一步需要知道在这个语言的种种不同的词里(包括词根、词头和词尾)，应当用什么样的区别性特征的小的矩阵来代表。最后一步，就是要写成一套规则，这些规则必须写得很严格，一般分三部分：输入部分，输出部分，条件部分。规则之间往往有固定的次序，不能任意颠倒。Chomsky 和 Halle 在《英语的音型》里还指出，有些

规则不仅有固定的次序，而且还有象环 (cycles) 一样的规则。也就是说，有一些规则要到环行的顺序中运用好几次才会显示出来。可见规则之间的次序有时候是相当复杂的。

〔附录二〕 关于声调语言

世界上的声调语言比一般人想象的要多得多，不限于中国或者东南亚。其实世界各地都有声调语言。比如，欧洲的挪威语、瑞典语、南斯拉夫的塞尔维亚语都是声调语言。非洲东部所有的班图语里，只有斯瓦希利语 (Swahili) 不是声调语言。美国和南美洲大部分印第安语都是声调语言。新几内亚和澳大利亚的很多语言也是声调语言。

但是，各种声调语言的性质并不相同。很多声调语言的声调只限于高、中、低的差别，没有升降曲折。比如，有人认为日语也是声调语言，因为日语声音高低的不同能区别词义。同是 はし (hashi)，前一个音节は读得高，意思是“桥”，后一个音节し读得高，意思是“筷子”。有时声调的高低要在它邻近的音节表现出来。比如はな (hana)，既指“花”，也指“鼻子”，都是第二个音节な读得比较高，听起来一样。但是，如果后面接着还有其他音节，比如后面接着系词です (desu)，这个系词在“花”的后面读得比较低，在“鼻子”后面就读得比较高。换句话说，两个“はな” (hana) 之间虽然有区别性特征，但是自己并不能区别开，一定要在后面的音节上(其他语言可能在前面音节上)才能表现出来。

十几年前我研究声调语言的时候，曾经到瑞典和墨西哥去了好几次，当时希望能把世界上的声调语言作一个归类。我把上面所说的这种只在本身之外表现声调的现象叫做“外部重音” (external accent)。在这方面，最令人感兴趣的是一种很少人知道的语言，在墨西哥，叫做 Mixtec，它的外部重音是可以连续使用的。我

曾经利用一次假期花了三天时间找到那个村庄，在那里住了两个星期。当地大约只有一百多人会说这种话。Mixtec 语一共有四个不同的声调，都是平的。除了高、中、低之外，还有一个“提高”(up)调。凡是有提高调的地方，就把下一个音节的音高再提高一步。这种提高调是可以连着使用的。有一次我做了一个每个音节都是提高调的句子请发音人念。他念了好几次都念不出来，因为整个句子的音节不断提高上升，念不了那么高。非洲还有一种语言，被称为“梯级语言”(terraced level language)，它的情况和 Mixtec 语完全相反，不是提高，而是往下降。

世界上声调语言虽然很多，但是象汉语这样每个音节都有固定声调，不但有高低之分，还有升降曲折之别，却是不多的，可以说这是我们特有的丰富的语言材料。研究汉语声调，一般都用赵元任先生几十年前介绍的五度制声调符号。用这套符号记音，是很有用的，但是，如果进一步分析其中哪些是有区别功能的，也许就不够用了，因为这套声调符号是只注意音值，不管声调的音位分析的。因此，常常出现这种情形：同一个方言的某个声调，有人说是〔55〕，有人说是〔44〕，又有人说是〔54〕或〔45〕。到底哪一个对呢？也许大家都对，因为可能在这个方言里这四种调值之间根本没有区别功能。

在美国，有很多人注意到汉语的声调问题，常常用声谱图把各种汉语方言的声调画出来，研究其中的问题。但是这个办法并不十分成功。可以举一个例子。普通话的上声在另一个上声之前究竟变成什么，看法很不一致。有人说变成了阳平。有人说和阳平不同，应该算是第五个声调。于是就有人认为，有了语图仪，看看它在窄带声谱图上是不是和阳平完全一样，就能很客观地解决这问题了。但是事情并不这样简单，因为声谱图也并不能解决一切问题。在声谱图上，不只是阳平和上声变出来的所谓“阳平”有时一

样，有时不一样，就是同一个人说同一个字，说两三次都可能不一样。既然不一样，就很难肯定变出来的究竟是不是阳平。所以不同的问题需要不同的实验方法来解决。

六十年代，我曾经和我的朋友李功普一起想出一种实验方法来解决这问题。① 我们没有依靠声谱图，而是找了许多对和这种变调有关的词和词组，例如：

A组：买马　有井……

B组：埋马　油井……

A组是上声加上声，B组是阳平加上声，一共找了一百三十多对。我们先请一位在北京住了很久的人来录音，录音的时候把A组和B组混杂在一起。然后找了三十多个能讲相当标准普通话的中国人来听录音。我们给他们一张写着这一百三十多对词的表格，排列次序也是A组和B组混杂的，请他们听到哪个词，就在表格上用铅笔画下来。测验的结果，没有一个人能答对55%以上。我们又请录音的人自己听。她能回答对的也没有超过60%（大约是58%）。也就是说，实际是无法区分A B这两组词的。回答只能对一半左右，说明是在那儿猜。既然是猜，就说明这里没有音位上的区别。所以，我们可以毫不犹豫地做出结论：在上声前面由上声变出来的声调就是阳平，不是什么第五个声调。

我们还可以从另外一个角度来说明这个问题。如果有人认为英语[p] 和 [b] 是一个音位，也就是认为 pin 和 bin 的声音一样，我们也可以用上面的实验方法证明他是错的。录下许多对 pin 和 bin 之类的词，按照上面的方法请人听录音回答，肯定每一个人都能回答对的。如果上声变出来的声调确实和阳平不是同一个音位，象英语的 [p] 和 [b] 那样，那么，每个人起码都应该能够答对

---

① W.S-Y. Wang and K.P. Li, 1967, «Tone 3 in Pekinese», Journal of Speech Hearing Research, 10.3. 629—36。

百分之九十以上。

1967 年，我写了一篇《声调音位特征》(Phonological Features of Tone)①。这篇文章提出了声调的区别性特征问题，虽然可能有不少错误，但是至少是提出了这个问题。我把声调分成下列七个区别性特征：1. 调形 (contour), 2. 升 (rising), 3. 降 (falling), 4. 曲折 (concave), 5. 高 (high), 6. 中 (mid), 7. 央 (central)。

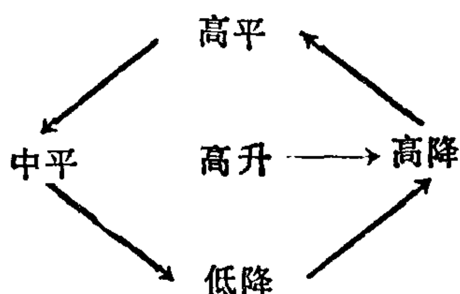
“调形”(contour) 是指平仄的分别。“曲折”(concave) 是指〔315〕或〔151〕之类的曲折调型。“高”(high) 和“中”(mid) 可以区分四种音高。有很多语言只有两种音高，就可以只用一个区别性特征来表示：高的叫“+high”，低的叫“−high”。如果不只两种音高，就必须再用上另一个区别性特征“中”(mid)，这样就可以分出四种音高：

	˥	˦	˨	˩
H	+	+	−	−
M	−	+	+	−

起初我以为这就够用了。后来有人告诉我，中国有些少数民族语言，比如一些苗语，就有五种不同音高。我只好又加了一个区别性特征“央”(central)，如果有五种音高，就用它表示最当中的一个。但是这种情况是比较少的。

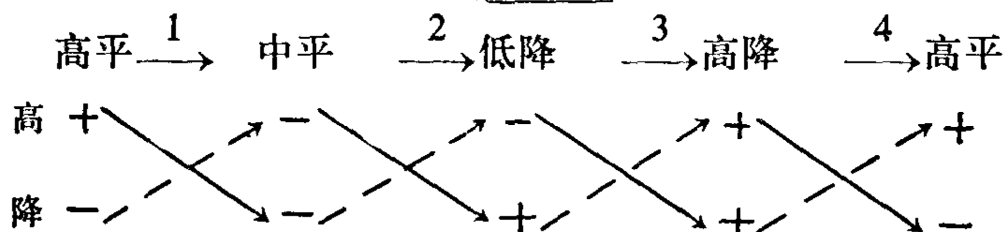
下面我想举一个例子来说明怎样用区别性特征来分析汉语方言里的声调。在我研究声调的区别性特征的时候，看到了一本书，这本书纪录了东南亚吉隆坡地方说的一种闽南话。这种闽南话和厦门话很相似，有两个短促的入声调，还有五个长的声调。据这本书说，这五个调子分别是高平、高降、低降、中平、高升。它们的变调情况相当复杂，可以用图表示如下页左上。这种声调变化象时

① 发表在 International Journal of American Linguistics 33.2.93—105。



钟走动一样，是按逆时针方向变化的。在闽南话里这一类情形是很不少的。我就用这个材料来分析它的区别性特征。

首先必须断定的是，在这个系统里哪些特征是必要的。我想，只有“高” (high) 和 “降”(falling) 这两个区别性特征是必要的。从这两个特征就能区别四个不同的声调。虽然这个方言一共有五个声调，但变出来的只有四个，高升调和低降调在一定条件下都变成高降调了，因此一共只需要两个区别性特征就够了。现在，把它们的变化排成下表：



这些变调现象也可以用一条规则把它们写在一起：

$$[\quad] \rightarrow \begin{bmatrix} \beta & \text{high} \\ -\alpha & \text{falling} \end{bmatrix} \Big/ \begin{bmatrix} \alpha & \text{high} \\ \beta & \text{falling} \end{bmatrix}$$

这条规则用不同的变数 (variable) 把相当复杂的情况归纳在一起了，后来有人就管它叫“α, β 成对变数” (α,β−paired variables)。

从这条规则可以看出“高” (high) 和“降” (falling) 这两个区别性特征之间的密切关系。我们可以写成如左，举出一个具体例子验证一下。

输出音段　　　　条件

$$\begin{bmatrix} +\text{high} \\ -\text{falling} \end{bmatrix} \Big/ \begin{bmatrix} +\text{high} \\ +\text{falling} \end{bmatrix}$$

条件里是＋high，变出来的falling一定是相反的符号，即－falling，因为在规则里正是 α变成－α 的。（这是异化现象，在前面矩阵里用实箭头表示。）条件里是 ＋falling，变出来的 high 一定是同样的符号，即＋high，因为在规则里 β 是仍旧变成 β 的。（这是同化现象，在前面矩阵里用虚箭头表示。）这个例子变出来的是哪一对呢？ ＋high，＋falling变成＋high，－falling，必然就是高降变成高平的那一对。由此类推，每一对变化都可以用这个规则来说明。

学 术 报 告

〔美〕王 士 元

一 语言的发生

语言的发生问题就是语言的起源问题。这个问题在西方的传统文化里很早就引起人们的兴趣了。《圣经》的开头部分就有好几个地方提到上帝是怎样把语言传给亚当的。因为基督教是犹太人创立的，所以直到今天还有人以为人类最早的语言就是希伯莱语。还有比《圣经》更早的记载。最有名的是一个古埃及皇帝的故事。他把两个刚生下的婴儿放在一个屋子里，不让他们听到任何人说话。等婴儿长大学话的时候，他就注意听，有一次听到发出来的一个声音象某种语言里的一个词，就由此推断人类最早的语言就是那种语言。[1] 这些自然都只是传说故事，但说明在国外很久以前就有人对这问题感兴趣。但是这些文献记载只限于西方文化，因此是片面的。中国古代文献里有多少这方面的记载，我不清楚。中国文化的发展和古希腊、古埃及、古希伯莱完全不同。如果国内有人能对这个问题做一些研究工作，我想，对人类的思想发展史是会有贡献的。

从埃及的故事到现在已经有两千年之久，这方面的记载非常之多，可是没有什么有科学价值的。比如有人认为语言起源于原始人劳动时所发出的声音，这就是所谓"唷，希呵"理论。有人认为原始人是从摹仿大自然的声音开始慢慢建立语言的，这就是所

① 参看岑麒祥:《语言学史概要》，科学出版社，1958 年，第 8 页。

谓"叮哨"理论。还有人认为语言是从摹仿动物的叫声开始的，这叫做"Bow—wow"理论①(Bow—wow 指狗叫声)。这些理论之所以没有价值，一个很明显的弱点是把科学和宗教的概念混在一起了。另一个基本弱点是完全用哲学方法来推测，而不是用科学方法去研究，去实践。各种说法都是不可能有反证的，因而也就没有被证明的可能，有时甚至连问题本身都是模糊的。两千年来，尤其是近几百年来，为了研究这个问题，耗费了不少哲学家和语言学家的精力，从而阻碍了语言学的进展。所以，1866 年法国语言协会订了一条规则：在巴黎法国语言协会开会的时候，不允许有人做有关语言起源的报告，他们的学报也不接受有关语言起源的文章，因为可以猜想那大概都是一些废话。此后，许多语言协会也都订了这样类似的规则。在这以后的一百多年里，关于语言起源的问题就没有展开什么讨论。

1976 年，美国纽约州的科学院主持召开了一次规模很大的会，讨论的题目是"语言和语音的起源和演变"。参加这个会的有语言学家、人类学家、心理学家、动物学家、考古学家、地理学家等好几千人。这个会很受外界重视，«纽约时报»还做了详细的报道。会后把会上讨论的文章汇集在一起，印出了一本将近一千页的厚书，内容涉及十几个不同的学科。②这个讨论会是近几年语言研究中最精彩的活动之一。今天我想谈的内容就和它有密切的关系。下面围绕三方面的问题谈：一、关于语言起源的时间背景，二、语言和语音的关系，三、发音器官和大脑。

在语言起源的时间背景中，最关键的一点是世界上生物的历

---

① L. Bloomfield: «Language»(语言论)，1933 年，第 6 页。(国内现有袁家骅等的译本，可以参看。)

② Stevan R. Harnad, Horst D. Steklis, and Jane Lancaster, editors, 1976, «Origin and Evolution of Language and Speech», (语言和言语的起源和演变), Annals of the New York Academy of Sciences, volume 280, 1976。

史到底有多久。一百多年前，生物学界对生物的起源问题作了一些基本研究，提出了两种原则上对立的意见：一种认为各种生物都有它个别的来源；另一种相反，例如达尔文，认为所有的生物都是由一种或几种不同的原始生物分化、演变而来的。如果要把这两种不同看法进行比较，就必须同时考虑生物进化历史的长短。要是生物的历史很短，看不出什么变化，那么语言的来源，种种生物的来源，就变得很神秘，就只能承认语言是神给我们的，或者还有其他神秘的来源。相反，要是认为生物有很长很长的进化历史，那么上面的第二种意见就容易被人接受。也就是说，生物是从一个很简单、很原始的系统，经过亿万年的演化，然后才达到现在这种奥妙复杂的境地的。而人的手，人的眼睛，以及语言等等，也都是经历了漫长的演变过程的。

生物的演化史到底有多长？根据地质学、天文学、古生物学等许多学科的综合资料来看，地球大致形成于四十六万万年以前。地球存在的早期，自然是不会有生物的。在达尔文的时代，生物学家认为最初的生物出现在五万万年以前。但是二十世纪五十年代，发现了一些极为微小的化石，这些化石确实是古生物所留下来的，经过研究，对生物的历史又有了进一步的认识。现在一般公认生物已经有三十六万万年的历史。在这样长的历史时期内，可以相信，有很多东西是可以产生，可以变化的。为了使大家看得更清楚一些，我们姑且把历史上的三十六万万年算做一年。一年有三百六十多天，那么，一天差不多等于一千万年。一天有二十四小时，那么，每小时等于四十二万年。这样推算下去，每分钟大约等于七千年。这就把时间范围大大地缩小了。在这个缩小了的时间范围内，如果元月一日生物开始出现，到十二月一日恐龙出现又死亡，十二月二十五日灵长目动物出现，十二月三十一日猿类出现，十二月三十一日晚上十一点周口店猿人才开始用火。由此可见，

生物的演化时间有多么长久。语言这么一个复杂奥妙的现象正是在这很长很长的生物历史背景中产生的。现在看到的一些原始文化遗迹都是在这一年的最后一分钟形成的，汉字也是在这最后一分钟出现的。但是，在这最后一分钟内发展的速度是过去无法比拟的。因为，从原则上讲，有了语言文字以后，文化演进的速度比生物的演进就要快得多。比如，生物从在地上爬变成能在天上飞，要用几千万年的时间；而文化上解决飞的问题，即使从意大利人达芬奇最早开始设想飞机算起，到进入现代喷气机的时代，也只用了不到一千年。语言的演变，有它的生物部分，也有它的文化部分。当语言的演变有了它的生物部分的基础以后，它的演变就跟着文化前进，速度就变得非常之快了。

对生物演变的时间背景有了了解以后，就可以进一步讨论这一段时期内生物之间的关系了。这方面的研究过去大都是利用化石，看化石上骨头的形状，什么骨头接什么骨头，等等。这些都是很粗的办法。现在不仅仅依靠人类学和考古学，还可以根据遗传学和生物化学来进行研究。每种生物的遗传基因都有自己的遗传密码，在适当的研究条件下，可以把一种动物的基因分离出来。比如，把马和驴的基因各自分开，然后又把它们的一部分基因合在一起，看看是不是能粘合起来。根据遗传基因能否粘合在一起，粘合部分有多少等等，就可以测定马和驴的亲缘关系如何。所有的生物都可以用这样的办法来研究它们亲缘关系的远近。此外，免疫学也为这方面的研究工作提供了一些技术。比如，研究血球对于化学因素的反应，反应相近的，亲缘关系就比较接近，反应差得远的，亲属关系就远。由于各种学科的共同努力，现在对世界上两百多万类生物之间的关系，已经了解得相当清楚了。

从灵长目动物开始（六千万年前），逐渐分出了各个支派。其中一支是新世界的猴子，即南美、北美的猴子，保存了较多的原始

特征。另一支是旧世界的猴子，原始特征比前一支少。大约在两千万年以前，才开始出现猿类。猿的分类，已经研究得相当清楚。有一组在亚洲，包括长臂猿和东南亚的一种猿。在非洲也有两种猿：一种是大猩猩，一种是黑猩猩。根据生物化学和遗传学的方法来研究，人和黑猩猩的关系是非常密切的。灵长目动物至少已经有六千万年的历史。人和黑猩猩都是从古猿演变来的，分开的时间只不过在一千万年以前。由于黑猩猩和人在生物演变的过程中有那么长的相同阶段，因此有很多共同的地方。例如，黑猩猩牙齿的种类和数目跟人类最相似，脑的结构和神经系统也跟人类最相似，血球的构造百分之九十九以上是一样的。因此有人说，人和黑猩猩的差别大部分只是量上的差别，而不是质上的差别。所以，要研究语言的起源和演变，一定要研究新世界的猿和猴子是怎样变成人的。因为到了这个阶段，才开始有了这样丰富的用于交际的工具——语言。①

从古代的猿变成现代的人，是和生活环境有密切关系的。猿猴生活在森林里，在树上活动，睡觉、走路是用四只脚。现在非洲有一种很大的猴子，也还是用四只脚走路的。后来，由于地球上气候发生了极大的变化，许多地区的森林逐渐消失。一部分古代的猿猴因为不能适应环境的变化而灭绝了；一部分找到了新的森林定居下来，进化为现代的猿类，也就是黑猩猩、大猩猩等等；另外一部分移居到平原，它们就是现代人类的远祖。②

这部分猿猴移居到平原以后，由于环境的变迁，生活方式发

---

① William S-Y. Wang(王士元), editor, 1982,«Human Communication: Readings on Language and its Psychobiological Bases»(人类交际：语言及其心理生理基础读本),San Francisco:W. H.Freeman Company。

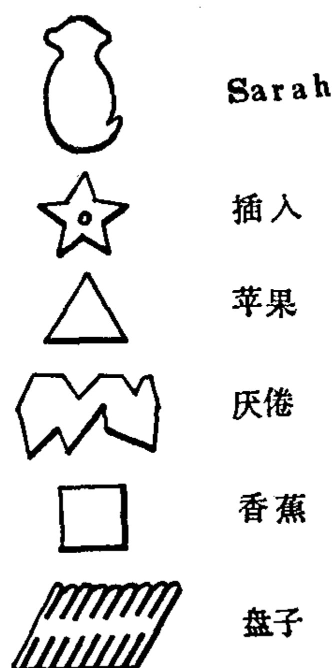
② 人类学家对这个问题的最新说法，请参看 D.C. Johanson 和 T.D.White «A Systematic Assessment of Early African hominids»(早期非洲人种的系统估计), Science 202—321—30,1979。

生了变化，于是逐渐从四只脚走路变成三只脚走路(现在我们看到的黑猩猩和大猩猩，它们走路就等于用三只脚，因为总有一只手按在地上当脚用)，躯干逐渐发展成半直立状态，最后完全摆脱了用上肢帮助走路的习惯，直立起来用两只脚走路。从身体与地面平行用四只脚走路到直起身体用两只脚走路，这是和语言的产生有密切关系的。用四只脚走路的时候，吃东西，移动东西，打架，各种动作都只能依靠嘴。嘴担负了这么多的职能，就不大可能发出各种不同的声音，语音产生的可能性也就小了。站起来以后，拿东西、打架等等动作都可以用手来代替了，嘴就逐渐变成了进行交际的重要工具。同时，由于站起来以后，喉头受地心吸力的影响渐渐往下移，从嘴到声门间的空腔慢慢扩大了，因此增加了发音能力。由于这些功能上的变化，人类的嘴的发展就和别的灵长类动物有了很大的区别。对于人类语言的产生和发展来说，这一步发展是有着极其重要的意义的。

从三十年代开始，美国很多学者在发现黑猩猩和人类有密切关系的基础上，猜想黑猩猩也许是可以说话的。一些心理学家在这方面进行了不少实验，其中比较有名的是 Keith 和 Cathy Hayes 所做的实验。他们在家里养了一只黑猩猩，叫 Viki，和自己的一个和 Viki 同岁的孩子放在一起照料。开始的时候，Viki 比他们的孩子能干多了，很早就会跑，能拿东西，抢着出门，非常淘气。对比之下，他们的孩子就只能躺在床上，活动能力很小。但一两年以后，情形就不同了。小孩子越来越聪明，懂事了，会说话了。但是，他们无论化多少精力，都无法使 Viki 说出话来。直到最后，Viki 所能听懂的话，仍不超过十个词。于是就有人下结论，认为黑猩猩是很愚蠢的，它没有用符号象征思想的能力。但是后来事实证明，这个结论是错误的。

六十年代，有一对夫妇叫 Allen Gardner 和 Beatrice Gardner,

他们认为，Viki 学话之所以失败，并不是因为它没有用符号象征思想的能力，而是因为在长期演化的过程中，人类的嘴和黑猩猩的嘴走上了两条根本不同的道路，要黑猩猩的嘴去做人类的嘴所能做的事，这要求可能是不合理的。尽管如此，黑猩猩的脑子也许是很发达的。于是，他们饲养了一只名叫 Washoe 的黑猩猩，它是在不到一岁时被从非洲带到他们家里去的。他们不是去教它说话，而是教它学习聋哑人的手势语。结果，Washoe 在短短几个月的时间内就学会了一百多个手势语。(而 Viki 用了好几年时间才学会了四个发音很含糊的词) 不仅如此，Washoe 还会创造性地使用手势语，用它表达一种新的意思。比如，用表示“脏”的手势表示它对狗的厌恶，不喜欢，用表示开门的手势表示要打开收音机(在英语里，这两个动作根本不是用一个词来表示的，“开(门)”是 open，“开(收音机)”是 turn on)。它甚至还会把两种手势结合起来表示一个新的意思，比如它想吃西瓜，就用表示“水果”和“水”的两种手势来指西瓜，意思是西瓜是含水的水果。可见黑猩猩并不只是死板地学一些手势，而是有它的创造性的。



从 Washoe 开始，大家对黑猩猩的学话能力有了进一步的认识。此后，又有一些实验室分别饲养了一些猩猩，它们学的“语言”各不相同。比如，有一只叫 Sarah 的猩猩，学话的办法是在铁板上放上不同的符号，表示不同的意思。左图就是 Sarah 所用的符号的一些例子。Sarah 不仅学会了命令句，而且还学会很多不同的句法结构，象“Sarah 如果想吃香蕉，那么 Sarah 应该把铅笔放在抽屉里”这样复杂的句子，它也可以看懂。还有一只叫 Lana 的黑猩

猩，是由电子计算机带大的。它会按电子计算机的按钮，让计算机给它开门，让计算机叫人来跟它玩。它还可以通过电子计算机听懂一些问题，并且作出回答。目前，还有一些人正在对灵长类的其他猿类进行类似的语言实验。

现在的问题是，象 Washoe 这样的猩猩到底算是学会了语言没有。这要看你对语言下的是什么样的定义。如果你认为语言必须是有声的，那么它就没有学会。如果你认为语言可以是任何一种用A来代表B的符号系统，那就可以说它已经学会了一种语言。事实上语言系统是可以用种种不同的符号来表示的，语音只是其中的一种符号，手势是另外一种符号。人类在演化过程中把语言和语音联系在一起，这种联系是无法教给别的动物的。但是，黑猩猩和大猩猩在演化过程中和人类有那么长的相同阶段，所以它们也应该有用符号象征思想的能力。以前，我们认为语言离不开语音，所以得出一个错误的结论，认为别的动物都不可能有语言。现在只能说，别的动物不可能有语音，但还是可以有很简单的语言。在各种符号中，语音是最理想、最有效的符号。人类选择了这种符号，才使语言能达到今天这样丰富发达的地步。黑猩猩和大猩猩只会用另一种符号，所以在语言的发展上远远赶不上人类。

人类选择语音作为符号并不是偶然的，这是可以从生理结构上作出解释的。过去几十年，有些人在这方面做过一些研究，发现人的声门的部位比别的灵长类动物低得多。最近美国 Brown 大学的语言学系主任 Philip Lieberman 在前人的基础上做了进一步的研究。他对比了狗、旧世界的猴子、黑猩猩和人的声门部位，发现狗的声门最高，会厌和软腭有很大的交错的地方。旧世界猴子的软腭和会厌交错的地方减少了，因为它们半直立以后，肺和气管的位置慢慢地往下移，会厌也就跟着下降了。黑猩猩的软腭和会厌就完全分开了。人类的声门部位下降更多，会厌和软腭之间相距

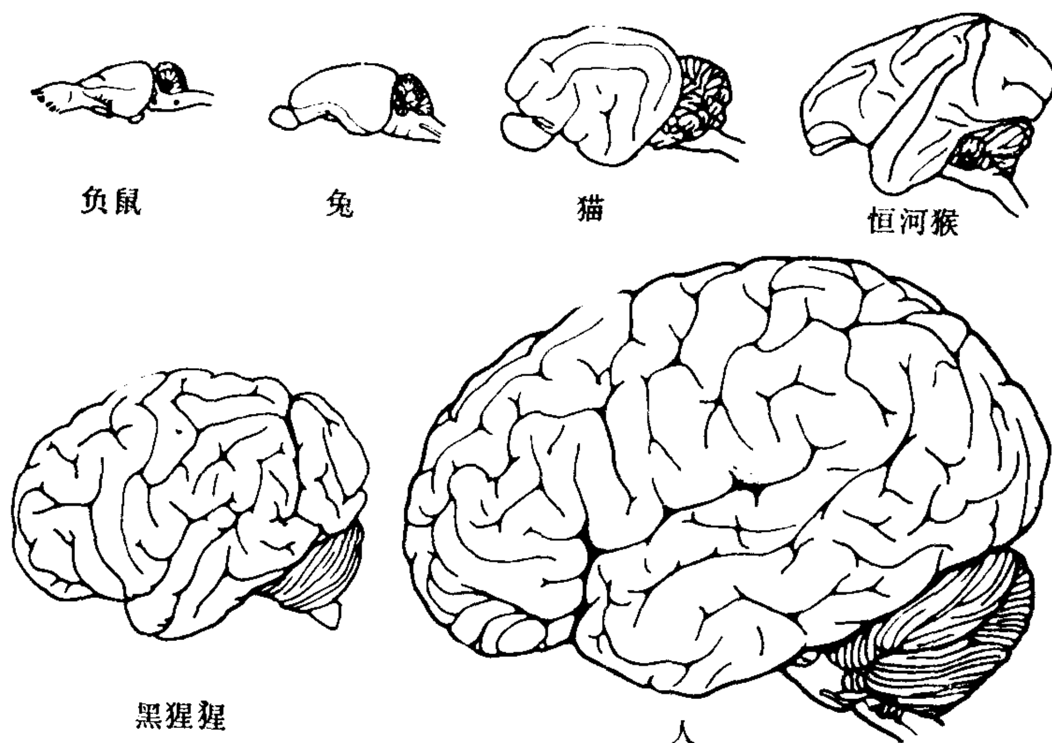
达几十毫米。这样，人的口腔和咽腔就形成了两个相当独立的气腔。两个气腔配合起来，就可以发出很多不同的声音。别的动物由于声门部位高，只有一个气腔，可以发出的声音就要少得多。Lieberman 的这些论述虽然有一些动物学家、生物学家和医学家表示不同意，但是我个人认为大致是可取的。声门下降确实可以大大增加音高控制的范围，也可以增强发音的能力。比如，常见的元音 [i]、[u]、[a] 都需要声门相当低才能发出来，[k]、[g]、[ŋ] 这一类舌根音也是这样，至于更根本的口音和鼻音的区别，更需要声门相当低才能分开。①

语音和声门的关系虽然这样密切，但是研究语言的产生，如果把注意力完全集中在这方面，方向就错了。因为人能说话，主要不在于多了几个元音和声调，而是在于表达思想的象征能力强，这种能力来自大脑，而不来自喉头。近十几年来，我们对于人的大脑，主要是大脑和语言的关系，有了不少新的了解。

把人和动物的大脑加以比较，至少有这几点重要的区别：第一，人的脑子比任何动物的脑子都大。大猩猩虽然身体比人大两倍到四倍，但它的平均脑量只有 400 多克，而人的平均脑量有 1,500 克左右。第二，人的大脑皮层的面积大。大脑皮层表面皱在一起，有许多非常紧密的皱纹使面积大大增加，脑神经的细胞可以达 140 亿个左右。第三，人脑发展的速度快。猩猩的脑子一生下来就相当完整了，脑子的重量差不多是最大的时候的一半。人类完全不同，初生婴儿的脑量只有 350 克左右，而且很不中用，各种神经之间的突触大部分还没有连接起来。但是以后它的发展速度很快，到两岁的时候，平均脑量是 1,000 克左右，到成年的时候，平均脑量是 1,400—1,500 克左右。由此可见，人生下来的时候，大脑只有四分

① Philip Lieberman, 1975, «On the Origins of Language»(语言的起源), New York, MacMillan Publishing Company。

之一的重量，在以后的发展过程中才逐渐完善起来，所以人受环境支配的程度比猩猩大得多。当然，脑量不能作为智力发展的唯一标准。但是，一般来讲，尤其是从动物之间的对比来讲，脑量的大小和智力还是很有关系的。下图是人和一些动物脑子大小的对比：



早在一百多年前，就有人发现大脑的左部和语言有很密切的关系。近几十年来，对大脑和语言的关系又有了进一步的认识。例如，哈佛大学的 Norman Geschwind 教授根据神经功能上的不同，把大脑皮层的语言中枢分成三个区域。[①] 不同的区域受伤后，表现出来的失语症各不相同。比如，如果其中一个区域受了损伤，听懂话虽然没有问题，但是说话时一个字一个字连不起来，特别是使用虚词发生了困难。如果另一个区域受了伤，表现就完全不同。

---

① Norman Geschwind, 1974, «Selected Papers on Language and the Brain» (语言和大脑文选), Boston: D. Reidel Publishing Company。

这种病人说话很流利，语调也正确，但是他说的完全是不合理的话，听起来不知所云，有些词甚至是他胡诌出来的。听人说话，看书也都会发生问题，因为他了解语言的机能出了故障。刚得这种失语症的人往往不知道自己有这种病，以为自己说的话都是有条理的，因此别人听不懂他就非常着急。在语言中枢的三个区域之间，还有一套神经把它们联系起来，如果这些神经损坏，失语症就表现为不能重复自己刚才说过的话。比如他说了一句:“你今天好吗?”你要他再说一遍，他虽然懂你的意思，但是做不到，因为联系三个语言区域的神经受伤了，不能把这个意思传递过去。

对语言起源问题的研究，现在越来越具体了。有一些人进一步考虑:脑子和语言的关系既然如此密切，这种关系究竟是后天环境造成的，还是先天就有的呢?这确实是一个很基本的问题。最近一些动物学家和心理学家对动物的行为产生了很大的兴趣。比如，有人发现小鸭子在刚刚出生后的十小时之内，如果给它看到一个正在动的东西，它就会老是跟着这个动的东西走，把它当作妈妈。象这种行为自然一大部分是先天的，不是教出来的。还有一个英国的生物学家以海鸥为对象进行了一项实验，几年前因此获得了诺贝尔奖金。他趁孵卵的海鸥外出觅食的时候，在海鸥蛋旁边放了一个木头做的非常大的假海鸥蛋，做得和真的完全一样。海鸥飞回窝来以后，判断不了哪个蛋是真的，最后竟然选择了那个特别大的木头假蛋来孵。海鸥的脑子告诉它，蛋越大越好。这也是一种先天性的行为。如果海鸥有后天形成的识别本领，那就很容易判断哪个蛋是真的了。

以上都是从动物身上观察到的现象:它们生下来的时候，脑子里就有了固定的电路，这些电路告诉它们在什么样的情况下应该有什么样的行动。那么，人类在学习语言的时候是否也有类似的情况呢?小孩学话的能力那么强，学习的速度那么快，是不是我们

生下来的时候，脑子里也有电路在指挥我们，让我们注意什么样的声音重要，什么样的声音不重要，以及这些声音之间的关系是什么样的呢？自然，这方面的问题是极其复杂的，还有待于进一步探索、研究。

现在，我们已经把一开始提出来的三方面的问题都谈到了。虽然对语言是怎么发生的这个基本问题目前还不能作出肯定的、完整的答复，但是，经过 1976 年纽约州的科学大会以后，这个问题的解决已经有了一个正确的方向。现在，在这方面进行研究的人不再是纸上谈兵了，不再是泛泛地空谈了，而是从各个方面把问题分成各别的、很小的题目，收集材料进行研究。有的问题已经得到了比较满意的答案，有的还在继续探讨。总之，关于语言起源的问题，现在已经提得比较科学、比较具体了，解决问题的基础也已经打下来了。问题不一定很快能解决，但是至少已经朝着正确的方向在一步一步地往前走了。语言是人类的一种最复杂、最奥妙的行为，牵涉到种种不同的活动。做学问象金字塔一样，基础广阔才能高。单单由一门学科，比如语言学、心理学或声学去研究语言，局限在一个窄小的领域内，收获一定小得多。要由许多相关的学科，不同的专家，一起协作，共同努力，才能获得巨大的成果。

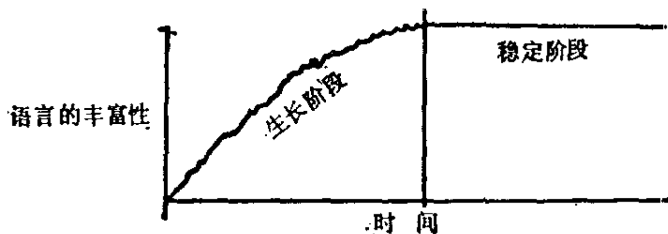
二 语言的演变

首先要说明的是，语言是和许多生理、心理现象紧密联系着的。有些人认为，语言这种机制在神经系统里是作为一个单独的部分存在的，和别的部分并没有很大的关系。有人甚至把语言看成是一种单独的器官，认为语言是人类演变历史中由于某种基因的巨大突变而产生的。N. Chomsky 在 1976 年讨论“语言和语音

的起源和演变"的大会上就曾经说过,语言等于是一种心理上的器官。这种看法未免把语言和语音的发生看得太神秘了。不过目前接受这种看法的人已经越来越少了。种种事实证明,语言并不是孤立存在的,它和许多生理和心理现象密切相关。

首先,语言和人的记忆力有直接的关系。人在老年或病重的时候,往往会忘记不少以前常用的语词。其次,它又和分析能力紧密相关。人可以把词组合成句子的这种高度的抽象分析能力,在各方面都有所表现。比如,音乐的结构和语言的结构是大同小异的,都是一层一层组织上去的,形成一种分层次的系统。现在已经有不少语言学家在注意语言和音乐在发生和演变上有没有互相联系的因素。最后,语言又和人的感觉能力有密切的关系。人耳所能听到的声音,幅度可以超过一百分贝,频率可以超过一万多赫。正是有了这样敏锐的感觉器官,才能产生象语音这样一整套复杂丰富的信号系统。这些能力——记忆能力、分析能力、感觉能力以及别的一些能力都是语言必须具备的,而且在语言发生之前必定已经存在了。别的动物,象黑猩猩,也有这些能力。举一个例子,非洲的黑猩猩爱吃一种白蚁,但蚁洞很小,手伸不进去,它就找一根又细又长的树枝伸到洞里,等白蚁爬到树枝上以后拉出来再吃。这说明黑猩猩也有相当强的分析问题的能力。人类的特殊发展在于能够把这几种能力联系在一起,综合成一种有系统的工具,这种工具就是语言的前身。语言发生以后,又对这些有关的能力起了反馈的作用,使它们发展得更快。别的灵长目动物则由于不能把这几种能力综合在一个系统里,在和人类的演变的竞争中越来越落后了。用电子计算机的术语说,语言是一种交接面,是把人类的有关智力的几部分综合在一起组成的。这种综合究竟用了多长时间才完成,这是一个现在还解答不了的问题。但是,从开始具备这种综合能力直到现在,这个漫长的发展过程,我想可以分为两大

阶段：生长阶段和稳定阶段。可以用下图来表示。①



如图所示，在开始阶段，语言的生长大概很不稳定，发展得对头就增加了语言的丰富性，发展得不对头就反而减少了丰富性。这样反复地经历一个很长时期，直到某一时候一个新的阶段开始，语言结构就比较稳定了。生长阶段到底有多长，还不清楚，也许曾经经历过几十万年甚至几百万年的时间。稳定阶段则是现在的情况。目前世界上各种语言都有很多共同的特征，即所谓"语言普遍性"(language universals)。有了这些特征之后，就意味着进入了语言的稳定阶段。这些特征把人类语言的范围固定住，限制住，语言的演变也就只能在一个较窄的范围内进行了。

人们对语言演变的研究是从什么时候开始的？这是一个属于学术史范围的问题。1968 年，U. Weinreich, W. Labov 和M. Herzog合写了一篇文章，叫做《语言演变原理的经验基础》(Empirical Foundations for a Theory of Language Change)②，这篇文章对语言演变做出了基本的贡献，研究历史语言的人不可不读。当然，他们的意见也并不是完全可以接受的。例如，他们认为这种研究开始于拉丁语的历史范围内，这就把语言学和历史语言学研究的开端只推到十六世纪初的意大利。这种说法显然太片面了，因为早在中国汉代，许慎在《说文解字·序》里就已经有很多地方提到语

① William S-Y. Wang(王士元)，1978，《The three scales of diachrony》(历时分析的三个尺度)，pp. 63—75 in 《Linguistics in the Seventies》, B. Kachru, editor, Department of Linguistics, University of Illinois。

② 载于《Directions for Historical Linguistics》(历史语言学的方向)，pp. 95—195, W. P. Lehmann and Y. Malkiel, editors, University of Texas Press。

言演变的问题了。中国语言学在思想史上是有很大贡献的，不幸往往被人忽略。希望国内有人在这方面多下功夫，把前人的贡献介绍出来。

在近代语言学里，对语言演变问题提出最为完整有力的理论的，还要算大约一百年以前德国的一个学派，就是所谓新语法学派(Neogrammarians，简写作 NG)。新语法学派在语言演变问题上有很多根本性的贡献，今天只介绍他们在语音方面的贡献。他们认为，要研究语音的变化，首先要明白语音的变化不是杂乱无章的，而是有规律的。这在当时是一个很有力量的理论。他们还认为，研究语音的变化，就是要找出这个规律，然后加以说明。所谓有规律，是指某个语音，比如说在二百多个词里头的某个语音，如果要变，就同时都变。他们认为语音变化可以从两个方面去看，一是作为词汇中的词，二是作为变化中的语音的不同特征。作为词汇，要变就都变，因而是一种突变。作为语音，变化是逐渐的，因而是一种渐变。这两者的关系正好相反。

把语音变化看成是一种渐变，这是由于一般研究语音的人观察不到语音总在变化。新语法学派认为，语音变化是逐渐的，变化的每一步都是很小的。这种看法从一百多年前提出来以后一直到现在，还是国外历史语言学中最为人们广泛接受的看法。① 丹麦著名语言学家 Jesperson 曾经作过一个比喻：就象锯木头一样，如果你想把每块木头都锯得一样长，每次都用锯下来的木头比着去锯下一块木头，只要稍不小心，开头和最后两块木头的长短就可能差得很远。Jesperson 认为，语音演变的情况正就是这样的：每次差一点，但积累几十年，几百年，差别就显著了。

---

① 参看《Resolving the Neogrammarian controversy》(解决新语法学派的争论)，William Labov，1981，《Language》57. 267—308。(会长致词全文，1979 美国语言学年会，洛杉矶)

但是这种说法仍然不免使人产生疑问。我在六十年代时就曾感到有些问题很难理解。举一个简单的例子：英语中音节开头的字母 k 曾经经历过一个历史音变过程，目前在一定条件下变得不发音了。比如 know，古英语发音是 [kn-]，现代英语是 [n-]（但如果 k 的前面还有元音，发音仍旧保留，如 acknowledge [əkn-]）。如果根据新语法学派的理论来看这个音变现象，就有问题了。这个 k 究竟是怎样渐变的呢？它总不能在这个词里变了百分之五十，在那个词里变了百分之六十五，又在另外一个词里变了百分之八十吧？k 要么就有，听得见；要么就没有，听不见。或是有，或是没有，不存在什么中间的道路。这样看来，变化就似乎应该是突然的，而不是逐渐的了。

但是，如果认为语音是突变的，也会产生问题。因为语音的变化的确不会是那么突然的。比如这个 k，绝不可能头一天晚上睡觉还在，第二天早晨就没有了。所以，音变总的来说仍然是一小步一小步走的。但在某些语言现象中，语音的变化既然是突然的，词汇的变化就不会也是突然的。这样看来，新语法学派的语言演变是渐变的理论，至少可以说是不全面的，不能用来说明所有的语言现象。

我把这种语音是突变、词汇是渐变的现象称之为"词汇扩散"（Lexical diffusion）。它和新语法学派语音渐变、词汇突变的说法正好相反（见下表）。

	词　汇	语　音
N G	突　变	渐　变
词汇扩散	渐　变	突　变

"词汇扩散"理论的基本内容是，我们虽然不容易看到语言在变，但很容易看到语言在任何时候都存在共时变异的现象，比如总是有些词具有两个或更多不同的念法。这种共时变异正是"词汇扩散"常常经过的途

词 \ 阶段	未变	变化中	已变
W_1			$\overline{W}_1$
W_2		$W_2 \sim \overline{W}_2$	
W_3		$W_3 \sim \overline{W}_3$	
W_4	W_4		
…			

径。[1]左表可以说明“词汇扩散”是怎样通过这种共时变异进行的。

W代表一个词，$\overline{W}$表示变化完成的词，$W \sim \overline{W}$表示变化中的词。在一般典型音变中，有些词变得较快，有些变得较慢。所以在音变过程中，总是可以把词分成未变、变化中、已变三类。比如表中 W_1已经完成了变化，W_2 和 W_3 有时是原有读音，有时是新的读音，W_4 则还没有变。

这里可以举一个具体的例子。英语中由双字母 oo 表示的元音 [uː] 在很多方言里已经在向 [ʊ] 变了， 也就是音长缩短了一些，舌位降低了一些。比如 look 读 [lʊk]，已经变了；room 有时读 [ruːm]，有时读 [rʊm]，两可；而 food 读 [fuːd]，没有变化。这就是“词汇扩散”的一种表现。这样来看音变过程就不再显得神秘了。

但是要使一个理论建立起来，就不能只有三五个例证，而是需要大量材料。六十年代，我们曾经以北京大学中文系《汉语方音字汇》* 为主，收集了现代汉语各方言以及中古音、日译吴音、日译汉音等近三十种方言的材料，每种都有两千多字音。我们用了两三年的时间把这些材料全部送进电子计算机去，[2] 现在用起这些材料来，既迅速，又可靠。这些材料为“词汇扩散”理论提供了大量例

① 《The Lexicon in Phonological Change》(音位转移中的词汇)，王士元编，1977 年，Mouton。几年来有很多用“词汇扩散”理论研究历史语言学的人，分析了十几种不同的语言，他们的部分论文收于此书。

* 编者按:《汉语方音字汇》现正由北京大学中文系现代汉语教研室修订。

② 共同做这个研究的有郑锦全、陈渊泉和谢信一等几位朋友。1973 年我回国时也做过报告，刊于《中国语言学报》1974 年第 2 期，1—26 页。

证。[1] 比如，汉语双峰话里古浊声母还没有全部清化，在音节开头的条件下，[+voiced]>[-voiced] 的音变，要受声调的支配，在不同调类中变化的频率不同：古浊声母在古入声字中全部清化，如“笛”[ˌt'i]；在古平上去声字中一般保留浊声母不变，如“狂”[ˌgɒŋ]，“病”[bin']，但少数也已清化，如“叛”[p'iɛ̃']。再如潮州话的声调，平上去入各分阴阳，共八个调。但近几十年已经有好几个人发现，潮州话二百多个阳去调字，已经有一百多个归入阳上调，如“健”['kieŋ]，另外一百来个还保持原来声调，如“阵”[tiŋ']。通常一个典型的音变大概是开始时慢，中间快，最后又慢。潮州话阳去调字变入阳上调的趋势大致已进行到一半，正好在中间。我们还可以举一个例子。广州话和香港粤语是非常接近的，但最近了解到，香港粤语鼻韵尾的发音部位有由舌根 [-ŋ] 移向舌尖 [-n] 的趋势，变化速度因韵腹元音舌位高低而不同：在高元音后变得快，在低元音后变得慢，在中元音后介乎二者之间。这是加州大学一个研究生研究的成果，发表在 1979 年的《中国语言学报》上。很明显，以上几个例子都不象新语法学派说的那样，所有的词汇都是突变的。从词汇上讲，它们其实是渐变的，几个词几个词地变过去。这正是“词汇扩散”的证据。

“词汇扩散”的现象不仅存在于汉语，也普遍见于其他语言。它不只是个别证据，而是和一般语言有关。比如英语在近几百年来产生一种语音变化现象：动词在用作名词的时候改变重音的位置。例如 affíx [ə'fiks]（动词）：áffix ['æfiks]（名词）。加州大学有个研究生，从古代不同时期的几部词典中去寻找这种重音改变的发展情况，发现 affíx 在 1612 年已经可以用作名词，但是重音

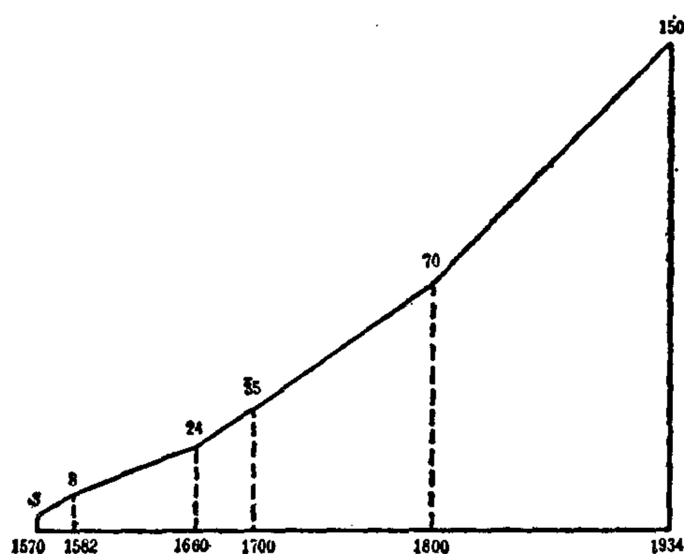
---

① William S-Y. Wang，（王士元），«Language change—a lexical perspective»（语言演变——词汇上的透视），1979 年，Annual Review of Anthropology, 9.353—71。

不变，重音的改变要到1775年的字典里才出现。情况大致如下：

词典年份	动词	名词
1533年	affíx	——
1612年	affíx	affíx
1755年	affíx	affíx
1775年	affíx	áffix

把类似 affix 的这类例证收集起来，观察这种现象在近三四百年历史中数量上的发展情况，可以归纳绘成下图。① 图中数字说明，1570年动词用作名词改变重音的只有三例，十二年后增加到八例，然后是二十四例，三十五例，七十例，直到1934年一百五十例，充分说明"词汇扩散"是随着时间一步一步增加变音词的数量的。



近几十年来，在汉语、藏语、英语、德语、法语以及印度南部方言等许多语言和方言里，都找到了大量的"词汇扩散"的例证，说明它是一种确实无疑而且普遍存在的语言现象。但是，一个新理论出现后，随着也会出现一些新问题。如果能解决这些问题，对于语言演变的认识就能更深入一步。

这些新问题中的第一个是，既然一种语音变化是有的词变得

---

① Donald Sherman, 1975,《Noun-verb stress alternation: an example of the lexical diffusion of sound change in English》(名动词的重音交替——英语语音演变中的词汇扩散一例),《Linguitics》, 149, 43—71。

快，有的词变得慢，那么，一个词的变化速度到底是由什么决定的呢？

很早就有人提出语音的变化和词的使用频率有关的说法，并且举出了几个例子来证明。但是，由于举的例子站不住脚，解释又欠客观，所以遭到很多人的反对。其实，当时所缺少的只是大量确凿的材料。几十年前，美国 Brown 大学的两位语言学家 Kucera 和 Frances 根据字典材料做了一个很好的、很可靠的英语词汇频率分析，从字典就可以看出哪些词常用，哪些不常用。最近，Joan Hooper 又在这个基础上前进了一步。他对“词汇扩散”的说法很有兴趣，收集了一百多个辅音后加 [əri] 的词，其中的 [ə] 都是非重读央元音 (Schwa)，把这样的词列成表格，请很多人念，并且问他们的语感，然后又把调查结果列成表格。表格上有一部分词，比如 nursery ['nə:sri]（托儿所）有时可以念成 ['n̩sri], armory ['a:məri]（军械库）有时可以念成 ['a:mri]。他发现，这个非重读元音 [ə] 失落的现象，是发生在使用频率高的词里的。但是他又发现，有些词使用频率虽然比较高，却并没有丢掉这个 [ə]。又经过进一步分析，才知道这是有特殊原因的，因为这些词如果要丢掉 [ə]，它前面的辅音和后面的 [r] 就会构成一个在英语里不能存在的辅音丛。因此结论仍旧应该是：一个词的频率越高，越常用，就变得越快；一个词的频率越低，越少用，就变得越慢。这是完全符合“词汇扩散”的理论的。①

语音变化可以分成典型音变和类推音变两类。后者在印欧系语言里特别常见，这是因为语法上的屈折性特别容易引起类推现

---

① Joan Hooper, 1977, «Word frequency in lexical diffusion and the source of morphophonological change» (词汇扩散中词的频率和形态音位变化的根源), pp.95—105 in «Current Progress in Historical Linguistics», W. Christie, editor, Amsterdam。

象。比如，英语中动词的过去时一般是用后加 -ed 的办法来表示的，有一些不规则常用动词则是以元音交替的办法来表示的。但是近来有一种趋势，不规则动词的过去时因为类推作用也逐渐变为后加 -ed。例如:

未变	变化中	已变
kept（保持）	crept, creeped（爬行）	bided（忍耐）
rose（起立）	leapt, leaped（跳跃）	reaped（收割）
knew（知道）	dove, dived（跳水）	mowed（割草）
lost（丢失）	dreamt, dreamed（做梦）	seethed（沸腾）

有趣的是，在类推音变中，变得较快的词不是使用频率较高的，而是较低的。词的使用频率和音变速度的关系，和典型音变恰恰相反。这种现象用“词汇扩散”的理论却是很容易说明的。Hooper 的看法是，这两种相反的趋向和儿童学话的情况可能有关。那些使用频率高的不规则动词，儿童在学话时很早就经常听到，就学过，熟悉得早，因此很早就固定下来，不再可能发生类推音变。典型音变恰恰相反，使用频率高的词听起来容易懂，容易懂的词一般说得比较快，说得快就比较容易丢掉其中某些不重要的语音成分(如非重读央元音 [ə])，儿童听到和学着说的机会比较多，就容易产生典型音变。

和“词汇扩散”有关的第二个新问题就是儿童学话的问题。语言演变和儿童学话是密切相关的，语言通过老一代传给年轻一代的时候，总是会有一些变化的。那么，“词汇扩散”和儿童学话到底存在什么关系呢?

深入分析儿童学话的过程，对了解语言的发展和进行语言的教学研究都是很关键的问题。现在有很多人在研究儿童是怎样从一岁、两岁直到五岁左右掌握如此奥妙的语言的。1941 年 R. Jakobson 在一篇叫做《儿童语言、失语症和语音普遍性》(Kinder-

sprache, Aphasie und allgemeine Lautgesetze）的文章中提出，儿童学话不但用音位的观点来对语言现象进行分析，而且还把每个音位分成区别性特征，儿童学习语言就是这样一步一步向前进的。这个说法非常有启发性。不过，由于当时材料不多，Jakobson 的这个理论现在看来是显得过于简单了，尽管它曾经有过很大贡献，但毕竟已经只是历史上的贡献了。现在根据一些实验证明，儿童学话并不是一个音位一个音位学的，更不是一个区别性特征一个区别性特征学的，而是一个词一个词学的，学会了一部分词，然后再慢慢地摸索出词和词之间的关系来。音位和区别性特征并不是儿童学习语音的基本单位。八、九年以前，谢信一曾经观察一个台湾儿童学习台湾闽语声母的情况。经过几个星期的记录，很明显地看出来这个儿童正是一个词一个词学的。一个词学会了，同一声母的第二个词又不会了；过了若干星期学会了这第二个词，又把第一个词忘了。这个儿童学话的时候，记忆变动的单位是词。而"词汇扩散"的最小单位也正是词。

这方面最完整最有系统的一项工作是语言学家 C.A.Ferguson 和他的一个学生在几年前进行的。① 他们调查了四个幼儿，从一岁多就开始记录他们的发音经历，一直记录到每个幼儿都学会五十个词为止。工作是做得非常周密的。下表可以说明他们调查的情况：

---

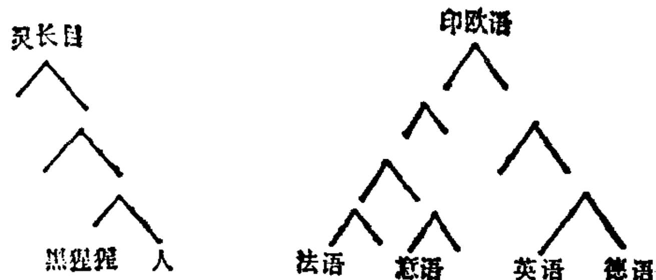
① Hsin-I. Hsieh（谢信一），1972，«Lexical diffusion: evidence from child language acquisition»（词汇扩散：儿童语言学习提供的证据），Glossa 6.89—104。Charles A. Ferguson and Carol B. Farwell, 1975, «Words and sounds in early language acquisition: English initial consonants in the first fifty words»（早期语言学习中的词语和读音：最初五十个词里的英语词首辅音），Language 51. 419—30。两文也见于 William S-Y. Wang（王士元）editor, 1977, «The Lexicon in Phonological Change», Mouton。

词例 \ 周次	六	七	八	九
baby	b~β	b~w~ph	b	b
book	b~Φ	b	b	b
bye-bye	b~ph	b~β	b	b
ball	b	b	β	b
banker	b	b	—	b
bounce	b	b	—	—
bang	—	b	—	—
box	—	b	—	b

从这个表里正可以看出幼儿学话的"词汇扩散"现象。他们从第六周到第九周学习双唇浊塞音 [b]，这个音在不同的词里进展的情况不同。在有的词里，开始的时候经常读错，但是很快就稳定下来，如baby, book; 有的词一开始学得很顺利，但是在中间偶然会糊涂一下，如 ball, banker; 有的词开头学得很顺利，但是过了两周就又忘了，如 bounce; 有的词学得一直不顺利，如 bang。我想这样的例子一定很多，很容易找，只是过去研究这方面问题的人受 Jakobson 的理论影响太深，不去注意。Jakobson 的理论可以和新语法学派的理论联系起来，而现在的这种看法则是和"词汇扩散"理论密切相关的。语言演变过程本来就是代代相传的，怎样去研究呢？恐怕只能以儿童学话的情况作为基础。

下面讨论第三个和词汇扩散有关的问题：语言或方言是怎样分类的？伟大的生物学家达尔文在一百多年前曾经说过，语言的演变和生物的演变相比，二者有很多平行的地方。这样，用来表示灵长目动物不同关系的树形图，也许可以同样用来表示语言或方言之间的不同关系。在达尔文的影响下，当时著名的语言学家 A. Schleicher 就把树形图借用到语言研究中去。下面是灵长目动物分类和语言分类的两种树形图。

在语言树形图中，语言分类是以音变为标准的。比如，甲、乙两方言浊声母都已经清化，就认为



它们的关系比较近；如果乙方言和丙方言开口韵有同样变化，就认为乙、丙两方言关系更近，等等。不过，能作为这种标准的音变材料不多。即使是历史材料如此丰富的汉语，从中古到现在的一千三百年里，目前也不过才有二十多种可靠的音变材料，其他语言要寻找这么多的材料就更困难了。可是，为了语言分类的可靠，这种材料自然是越丰富越好。谢信一在«中国语言学报»第一期里提出，不以音变作为语言分类的标准，而以音变影响所及的词作为这种标准。这样，标准就可以由几十个音变扩大为成百成千的词。他把江苏省四十几个点的方言材料（包括吴方言和北方方言）用“词汇扩散”的方法进行分类。尽管他研究的范围很小，材料也不很多，没有能引起人们的注意，但是他的研究结果表明，用“词汇扩散”的方法研究方言分类是很有希望的。①

Krishnamurti 曾经对印度南部的泰卢固语 (Telugu) 的方言群进行过类似的研究工作。他在这方面的研究工作和我们研究汉语方言有相似的地方。我们的出发点是中古音«切韵»音系的摄、等、开合口等等，他用的是一个化了几十年工夫做出来的叫做 DED (Dravidian Etymological Dictionary) 的材料，所以他可以知道每个词的词源是什么。② 泰卢固语是印度南部一个很大的方言群，其

---

① Hsin-I.Hsieh（谢信一），1973，«A new method of dialect subgrouping»（方言分类的新方法），«中国语言学报»1.64—92。

② Bh. Krishnamurti，1977，«Areal and lexical diffusion of sound change: evidence from Dravidian»（语音变化的地区和词汇扩散：达罗毗荼语提供的证据），«Language»54.1—20。

中各方言都经过一些不同的音变。他把各种方言中的音变现象加以分类，从中发现有“词汇扩散”的现象，然后用电子计算机根据词汇中语音已变、变化中、未变的三类词的数量计算，比较各方言间的关系，得出结果。他这个最新的研究结果还没有正式发表。这样的分类，我想在方法论上是有很大贡献的。几百个或是几千个词在“词汇扩散”程序中有不同的分布，把这个分布作为方言归类的基础，就有了比较客观的程序，而不象过去那样主要决定于个人的主观意见了。

目前一般人好象认为把语言树形图作为历史分析的工具是毫无问题的，是有效的。对这种看法我是怀疑的。我认为它过去之所以显得有效，是因为可供研究的语言材料太少了。当然，现在这样的树形图也还有它的用处，但是，从长远来看，等到材料逐渐丰富严密起来，语言树形图就未必还是合适的。在这方面，我和 L. Cavalli-Sforza 曾在南太平洋的 Micronesia 地区，以地理位置为条件作过一个调查研究。① 树形图的方法本来是从生物学借过来的。可是生物的代代相传和语言的代代相传是根本不同的。生物传代是由基因一代一代传下去，每一代生物有它的基因群或基因部，不同基因群的动物，比如狗和马，就不能生出什么“狗马”来。可是语言就不是这样。例如，两千年前一部分说日尔曼语的人侵入英格兰，和当地人密切接触，产生了英语。一千多年前，法国诺曼地人又侵入英国，统治英国一二百年，很多法语词在这时期进入英语。所以英语实际上不是一个单纯的语言。这种现象在生物学上就是不可能有的，在语言树形图里也不可能得到反映。这是语言历史中非常少见而又非常根本的问题，不过目前似乎还没有办法进一

① L. L. Cavalli-Sforza 和 M. W. Feldman,1981,«Cultural Transmission and Evolution: a Quantitative Approach»(文化的交流和发展: 一种定量的方法), Princeton University Press。

步去了解它。

语言演变问题是人类行为的研究中最为复杂的问题。因为它不仅仅是语言内部的问题，而且还牵涉到社会中人和人交往这个因素，又和人生理方面的发音、听觉、记忆等方面有关。所以，虽然对这个问题的研究已经有了几百年的历史，可是对它的了解到现在为止还仍然是很不完全的。要得到全面的了解，还要作出很大的努力。在这方面，我想中国语言学一定会起中心的作用，会作出重要的贡献。因为世界上再没有任何地方有中国那样丰富的历史语言材料和方言材料。无论是空间上还是时间上，中国的语言材料都是最为丰富的。而且，中国还有语言研究的历史传统，从许慎《说文解字》和扬雄《方言》开始，差不多每个朝代都有注意语言研究的学者。特别是近代的学者，对中古音和上古音的研究都有根本性的贡献。有人说，西洋的科学方法是在化学、物理的实验中形成起来的，而中国的科学方法是在差不多同时期——明末清初——在书本中，特别是在古书中探索出来的。我觉得这个对比也许有它的道理。这样悠久的传统不能不继承下来。当然，目前研究语音演变，不能只注意书本了，一定还要注意许多别的方面，特别是要利用电子计算机、实验语音学等现代技术。也需要改变汉语研究和少数民族语言研究互不相关的局面。现在学术性刊物《方言》《民族语文》创刊了，这是一个很好的发端，给我们增加了一支生力军。目前，语言演变的一些理论问题还不能解决，一方面是因为缺乏足够质量和数量的语言材料，另一方面也是因为在学术思想上太受印欧系语言系统的束缚。在这方面，我希望能在中国语言学的范围内得到一些新的了解，而这也是国外很多人的希望。

三　语言和文字的生理基础

我们现在大概都相信语音跟语言的发生和发展是有密切关系的。虽然语言的符号不限于声波上的信号，但是正因为人类选择了语音作为符号，我们的语言才能发展得如此完善、丰富。可是语音到底有它的局限性，最大的局限就是它的暂时性，一发即逝。这使语言在时间上和空间上都受到了限制。因此，文字的发明在人类历史上是非常重要的一件事。今天，我想借这个机会讨论一下语音和文字的关系，介绍最近研究文字的一些新收获。

和语言的起源相比，文字的产生要晚得多。通常认为语言至少有几十万年，甚至一百万年的历史。可是文字的产生，照目前的材料看，大概不会早于一万年以前。文字的发明可能和农业的产生有关。根据目前的研究，农业的发明差不多是在一万年以前。

语言的产生比文字早得多，文字是从语音中发展出来的。但是，不同语言发展出来的文字，所走的道路是不同的。各种文字跟语音的关系和距离是不一样的。根据这种不同的关系和距离，也许可以把世界上的文字分成四大类。

和语音关系最密切、距离最近的文字大概要算朝鲜的谚文。朝鲜的谚文是一笔一笔写的，和朝鲜语语音里的区别性特征很有关系。写起来，元音是一个系统，辅音是一个系统。在元音里怎样写是前元音，怎样写是后元音，在辅音里怎样写是送气，怎样写是不送气，这些在谚文里都能表达出来。从各方面看，朝鲜文和语音的关系都是最密切的。

在印欧系语言里，比如英语、俄语、德语、西班牙语等等，文字单位所代表的是音段(segment)，而不是区别性特征。当然，不同的

文字和所代表的音段的关系也并不相同。比如英文字母所代表的音段，大部分不是音位(phoneme)，而是词音位(Morphophoneme)。sane 这个词里的 a 代表的是元音 [ei]，sanity 里的a代表的是元音[æ]，同一个字母代表了两个不同的音位，这是因为要保持词、词位和语法之间的简单关系，所以读音变了，音位变了，文字并没有变。又如 electric，如果后边加上 -ity 成为 electricity，最后的 c 就由[k]变成[s]，但是，为了保持词形的联系，文字上并没有作相应的改变。由此可见，以音段为单位的各种文字还是有差别的，有的离音位近一点，有的离词音位近一点。英文字母在很多场合代表的不是音位音段，而是词音位音段。

朝鲜文字代表的语音单位最小，英文、俄文这一类文字所代表的语音单位就比较大，至于日本文字——假名所代表的语音单位就更大了。假名所代表的是音节，例如かな(假名)两个字母代表两个音节 kana，每个音节包含两个音段。日文的假名是典型的音节文字。

在各种文字中离语音最远的要算汉字了。朝鲜谚文、英文、日本假名和汉字分别代表了世界上文字的四大类。无论从哪一种语言来看，文字和语音总是有一种固定关系的，只是这种关系的远近不同，距离不同而已。

文字虽然是从语音中发展出来的，但是有的时候也可以反过来影响语音的发展。瑞典语就曾经出现过这种情况。瑞典文象英文一样，也是一种音段文字。早在二三百年以前就有很多人指出，瑞典语里紧接着元音后边的某些辅音是不发音的。例如zöd(铜)，ved(木头)，med(同，连词)，这些词在拼写上有字母 d，但是不发音。[①] (这种音变现象，印欧语系许多语言都有，法语尤其典型，很

① 参看 Tore Janson 的文章，载于«Lexicon in Phonological Change»(语音变化中的词汇)。

多拼在元音后边的辅音都是有字无音。）到了二十世纪二十年代，瑞典的扫盲运动搞得很好，大批大批的文盲都能识字看书了。他们按字母读音，把元音后边的字母 d 也读了出来。于是，由于文字的影响，二百年来的历史音变又倒转回去了。我想，文字对语音的影响大概不限于音段文字，汉字一定也会有这一类的例子。汉字简化的时候把“賓、鄰、進”简化成“宾、邻、进”，繁体字的声旁是舌尖鼻音韵尾[-n]，简体字声旁是舌根鼻音韵尾[-ŋ]。说不定将来就有可能因为文字的影响，“宾、邻、进”的读音由舌尖鼻音韵尾变成了舌根鼻音韵尾。

汉字虽然比别的文字离语音远得多，但是也有它的好处。中国历史悠久，疆域辽阔。在语言的发展过程中，各地语音的变化很大，但是不论哪个方言区的人，都可以看同样的书报。从历史方面说，汉字的这个优点也是显著的。虽然很多汉字现在简化了，在读古书时出现了一些新问题，但是，教我们的一个中学生看懂两千年前的古文，仍然要比教一个美国中学生看懂几百年前的英文容易一些。象中国这样一个地广人多的国家，文化上能够如此高度一致，这在世界上是少有的。在这方面，汉字是起了一定作用的。相比之下，中国的近邻印度情况就复杂得多。印度也是一个人口众多、地域广大的国家，但是文字的情况要复杂得多。由于文字上的不一致，文化的传统、社会的习俗都受到很大的影响。别的且不说，就以一张五个卢比的钞票来说，票面上除了必须印上英文和印地文(Hindi)以外，还要印上其他十几种文字。如果不把这些文字都印出来，有时候买东西都成问题。

当然，也必须承认汉字也有很多不方便的地方。比如，上万个汉字要编成一个固定的次序就很难，打字、印刷、电讯、编排索引、计算机的输入和输出等等方面，至少在目前的技术条件下，汉字都给我们带来许多麻烦。不过，我们应该看到技术总是在不断进

步的，也许再过十几年、几十年，情况就会很不一样。①

文字是每个文明社会必不可少的基本工具，成千上万的人每天在学习它，使用它。因此，有关文字的研究工作如果做得好，对社会的进步就会有很大贡献。文字的好坏是一个特殊的问题，可以用科学的方法去研究它，了解它。比如，一个一个的字是怎样学会的，学会了以后，一个一个的字是怎样读的，这些都是语言行为上的基本问题，应当用科学的方法去研究它。当然，回答了这些应用语言学上的问题之后，并不是马上就可以根据它确定一个理想的政策，因为政策的制定并不只是一个科学知识的问题。但是，我想，有关语言文字的政策总是应当有最充分的科学基础的。

文字的单位是字，可以是字母，也可以是汉字。认字的目的就是要知道字所代表的意义。世界上一般语言所用的文字，字和意义之间都不直接发生联系，而是通过语音，然后才达到意义。汉字是否也是这样的呢？这个问题引起国外许多学者的很大兴趣。现在很多人都在研究这问题。目前国外有不少学者认为，汉字不需要经过语音这个阶段，可以直接从字形达到意义。

早在几十年前，有一些国家，比如美国、英国、法国，发现有很多小孩说话毫无问题，但是在认字方面发生了很大困难。有的小孩到了十几岁还是不能看书，阅读。这是一种病症，严重的叫“失读症”(Alexia)，比较轻的叫“诵读困难症”(Dyslexia)。在这些国家里，这个问题已经成为教育界的一个严重问题。1968 年，有一个叫 K. Makita 的日本人在《美国行为精神病学学报》(American Journal of Orthopsychiatry)上发表了一篇文章，题目是《日本儿童丧失阅读能力的实际情况》(The Rarity of Reading Disability in

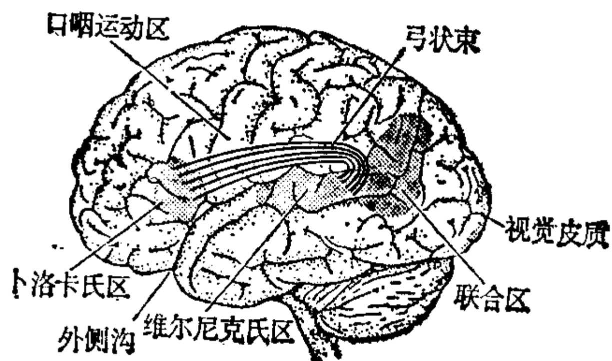
① W. S-Y. Wang (王士元)，1981，《Language Structure and Optimal Orthography》(语言结构和理想正字法)，The Perception of Print:《Reading Research in Experimental Psychology》，Tzeng and Singer，eds. Lawrence Erlbaum Assoc。

Japanese Children)。他认为，在日本由于学的是汉字和假名，所以日本儿童根本就不存在这样的问题。这篇文章引起了一些争论，有人认为这种差别可能是教育制度不同造成的，而不是因为文字的性质不同。但是差别确实存在，问题在于究竟是什么原因产生了这种差别。两三年以后，心理学家 P. Rozin 在 1971 年 3 月的美国《科学》(Science)杂志上发表了一篇文章，题目是《有阅读问题的美国儿童容易学会阅读用汉字表达的英语》(American Children with Reading Problems Can Easily Learn to Read English Represented by Chinese Characters)。他在美国东海岸的一所小学里做了一个实验，这所小学患有失读症的儿童比较多。实验的办法是用汉字代替英文，但不改动英语的语音，让患有失读症的儿童读。比如写出来是汉字"妈妈 放书"，读出来仍旧是 "The mother put the book"，也就是文字和语音搬了家。用汉字来解决失读症问题，实验的效果非常好，原来连看很简单的书都有困难的孩子，在改用汉字阅读以后，学得非常之快。Rozin 认为，小孩阅读英文字母拼的词的时候，大都是一个字母一个字母地拼出来，这对有失读症的儿童是很困难的。改用汉字就没有问题了，因为汉字是不怎么依赖语音的，它可以由字形直接达到意义。

1973 年，我有了一个小女孩。我觉得如果她长大以后根本不懂汉语，那是很不应该的事情。学习汉语并不难，最困难还是汉字。于是，我仿效 Rozin 的办法，在她两岁多的时候，就教她认汉字。我在一块块的小木头上用毛笔写上汉字，例如"妈妈、飞机、桌子、椅子、牛奶"，等等，但是教她用英语来读，"妈妈"读成 mother，"桌子"读成 table，"牛奶"读成 milk，等等。结果她在两岁零三个月到九个月这半年的时间内，在一个完全是英语的环境里，居然学会了一百多个汉字。而一般美国儿童在这么大的时候还没有开始认字母呢。当时她以为英语就是那么写的。一直到四岁多进

了幼儿园，才发现英语原来不是那么写的。从我女儿学汉字的例子来看,汉字未必象许多人说的那么难学难认,问题也许还是在于学习的时间、学习的方法和学习的环境。如果把这一类比较细致的问题都弄清楚了,都解决了,也许学习汉字并不是什么很麻烦的事情。汉字毕竟还是有很多基本优点的。

世界上一般语言用的是拼音文字:是由字母读出语音,然后再由音达到意义。许多儿童看书看得慢,是因为他并不只是在看字,他还在默默地拼,喉头的动作自然比眼看慢得多。大部分人在认字的时候都经过这个阶段,先通过语音,再得到它的意义。汉字可能并不是如此。它不需要经过语音,可以从字直接达到意义。Makita 和 Rozin 的两篇文章在这方面有很大的启发性。但是我们还需要进一步的科学的证明。近几年来,出现了一批很有趣的文章,主要是在日本。日本很重视医学和语言学的关系,语言学家常常和医院合作进行科学研究。最值得介绍的是一位日本女学者 Sumiko Sasanuma,1974 年她在《中国语言学报》(Journal of Chinese Linguistics)2, 141—158, 1974 年上发表了一篇文章,题目是《日本失语症患者在书面语言上的障碍: 假名和汉字的处理》(Impairment of Written Language in Japanese Aphasias: Kana versus Kanji Processing)。她研究的主要是一种失语症(aphasia)。这种病症常常是大脑受伤引起的,比如脑溢血、摔跤、撞车、殴斗等都会引起失语症。Sasanuma 就从这种失语症来研究文字是怎样和语音、意义联系的。



我们知道，语言主要是由左大脑掌管的。左大脑的分区如右图。Sasanuma把卜洛卡氏区 (Broca's area) 受伤的病人叫第一类病人，维尔尼克氏区 (Wer-

nicke's area) 受伤的叫第二类病人，这两类病人在文字阅读上的表现是不同的。例如，第一类病人里有一个受过高等教育的43岁的工程师。他到医院的时候已经不怎么能说话了，但是还能写字。他在写字的时候，有一种很有意思的现象。这个工程师在写汉字的时候一点问题也没有，（一般日本人至少认得两千多个汉字）可是假名就几乎完全忘掉了。比如请他把“女の人が手紙を書いている”（那个女人在写信）这样一句话写出来，他就写成“女の人 手紙書”。这在日文里是完全不通的。他单把汉字写了出来，假名除了一个“の”以外都不会写了。Sasanuma 接着做了进一步的实验。让他听写单词，比如“子供”（小孩），“着物”（和服），“封筒”（信封），这些汉字他写起来一点问题也没有。可是一听写假名，他不是忘了，就是写错了。其实假名笔划简单，比汉字容易写得多。从他写错的假名里可以看到，他的错误是有系统的。比如把ど[do] 写成了の[no]，大部分错误都在某个区别性特征上，或是清浊不对，或是发音部位不对。因此，可以说这个病人在假名书写上的错误实际上反映了语音判断上的错误，这和一般人写错假名的情况不同。Sasanuma 认为，这个实验说明，第一类失语症病人是左大脑管语音的部分受了伤。虽然假名写起来简单，但是因为它和语音关系近，所以病人不会写；汉字形状虽然复杂，可是和语音的关系远，所以倒都记得。

第二类失语症病人的情况恰恰相反。比如，有一个病人是29岁的中学毕业生。叫他写假名的时候，不管是片假名还是平假名，他都能写。但是写汉字的时候，就有很多写不出来，也认不出来，再不然就是写错了。他写错的时候也和一般人的错误不一样。他记得字音，但忘了字形。例如让他写“金钱”（きんせん kinsen），他就写成了“近线”。声音相同，但是写出来的字毫无意义，连他自己看了也不懂。更有意思的是第二类失语症病人把字的音读（由

汉字读音发展而来的）和训读（日语固有的读音）完全给弄混了。例如"田舍"，只有训读いなか(inaka)，没有音读，但是他照音读念成でんしゃ(densha)了；"中风"在日语里一般只用音读ちゅうふう(chūfu)，但是他又照训读念成了なかかぜ(naka kaze)。由于他把字的音读和训读弄混了，所以念出来以后自己也不懂是什么意思。象上面这一类很有趣的病例，最近八九年在日本收集了很多。这些材料有很大的启发性。很显然，文字的性质不同，在大脑里处理的部位也不同。在日本的这些病人当中，左脑前部受了伤，书写假名就不行了；左脑后部受了伤，书写汉字就会发生困难。

这方面的研究成果虽然富有启发性，但是从科学观点来看，还是有它的局限性的。因为首先，这种失语症病人毕竟只是少数，不能只根据少数病人的情况就得出一般性的结论。其次，这些失语症病人虽然脑子这一部分受了伤，但是我们不知道是不是还有别的部分也受了伤，因为如果别的部分受了伤，也有可能会影响到这一部分的功能。要进一步解决这方面的问题，就必须从健康人的身上去寻找答案，一步一步地去研究。这方面的问题是很多的。下面只想谈两个问题：一个是汉字是不是真的和语音没有关系，另一个是汉字和大脑的关系。对这两个问题的研究，我看最有贡献的是一个年青的心理学家，叫曾志朗(Ovid J.L.Tzeng)。下面就介绍他的一些研究成果。

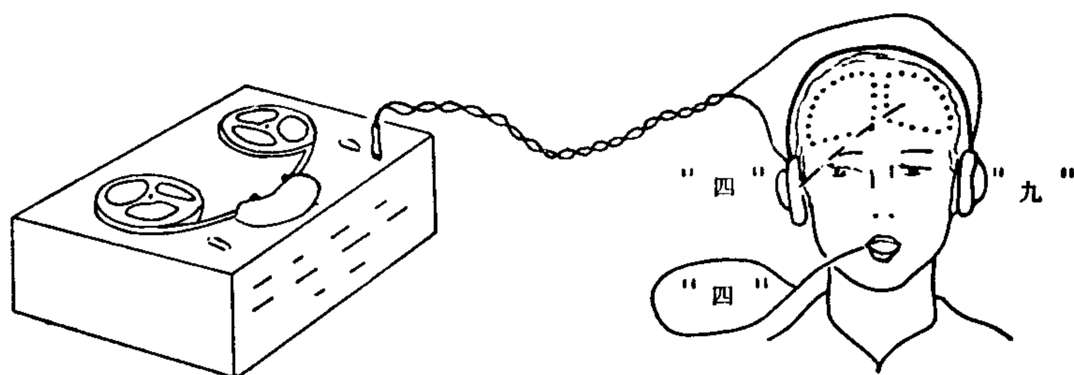
前面说过，有些人认为，汉字并不通过语音，而是由字形直接达到意义的。这种说法究竟在多大程度上是符合实际情况的呢？就这个问题，曾志朗做了一个很有价值的实验，发表在1977年的《实验心理学学报》(Journal of Experimental Psychology)3、621—30上，题目是《阅读汉字时言语的再编码》(Speech Recoding in Reading Chinese Characters)。他首先选择三组语音性质不同的汉字，让一些中国人参加这个实验：

第一组：SC(同辅音的字)。例如：克，康，开，哭，口……

第二组：SV(同元音的字)。例如：七，吉，西，李，你……

第三组：SCSV(同辅音、同元音的字)。例如：石，示，十，事，史……

实验分三个阶段。第一阶段是认字。用幻灯机放四个字，一秒钟一个字。四个字都是从同一组里选出来的，或者全是SC，或者全是SV，也可以全是SCSV。要求看的人记住幻灯片上的四个字。第二阶段是干扰(interference)。给他带上耳机，放送预先在录音带上录制的六个字，要求他一边听，一边跟着念。(见下图)一共干扰他



两次，每次六个字。这些干扰的字也是同一组取出来的，可能和第一阶段看的字同组，也可能不同组。例如看的字是SV，干扰的字可能是SV，也可能不是SV，而是SC或SCSV，但是各次干扰必须是同组的字。第三阶段是测验记忆。要求他把第一阶段看到的字按照原来的顺序写下来。我们知道，第一阶段是完全不发声的，全凭视觉看字；第二阶段完全靠语音，全凭听觉念字。如果汉字真的和语音完全没有关系，第二阶段就不会干扰第一阶段，也就不会影响第三阶段的记忆。但是实际上并不如此。下表是对几十个人测验的结果：

干扰字表 (interference list)	测验字表 (Target list)			
	SC	SV	SCSV	平均
SC	0.47	0.70	0.55	0.57
SV	0.71	0.35	0.62	0.56
SCSV	0.66	0.51	0.32	0.50
平　　均	0.61	0.52	0.50	——

表中的数据，是根据几十个人的测验结果统计出来的。测验字表中第一组字 SC，受 SC 干扰以后，到第三阶段只能记忆 0.47，能记住的不到一半；受 SV 干扰以后，能记忆 0.71；受 SCSV 干扰以后，能记忆 0.66；平均能记忆 0.61。从这一组的情况来看，当干扰字和测验字是同组(SC-SC)的时候，干扰最大；当干扰字是 SV 的时候，干扰最小，从 0.47 增加到 0.71；SCSV 的干扰则介乎二者之间。其他两组测验字的情况也是如此。比如第二组测验字 SV，受 SC 干扰，能记忆 0.70；受同组字 SV 干扰，下降到 0.35。受干扰最厉害的是第三组测验字 SCSV，它受同组字SCSV 干扰时，只能记忆 0.32，也就是说，几百个字当中能记住的不到三分之一。从表上还可以看到元音的干扰性比辅音强。这和几十年来研究语音听觉方面的实验结果是一致的。在大脑里，元音的影响通常是要比辅音强得多的。

上面的实验是以单个汉字作为实验材料的。但是我们看书的时候并不是一个字一个字看，而是一整句一整句看的。那么，在一整句一整句看的时候，到底和语音有没有关系呢？就这个问题，曾志朗又作了第二个实验。他以四种句子作材料。第一种，押韵，而且成句，例如“糊涂夫妇砍树木”。除了一个“砍”字，其他字的韵母都是[u]，也可以说都是押韵的。也就是说，句子里字和字之间有语音干扰。第二种，不押韵，但是成句，例如“迷糊夫妻摘花草”。不押韵就是字和字之间没有语音干扰。第三种，押韵，但不成句，例如“糊树涂夫砍木妇”。第四种，不押韵，也不成句。把这四种句子分别用幻灯片放出来让人看。如果认为幻灯片上的字成句，就按一个电钮，如果认为不成句，就按另一个电钮。要求按得越快越好。

这个实验和前一个实验的道理基本上是一样的。判断幻灯片上的字成句还是不成句，只需要知道意义就行了。如果从字到意义是不经过语音的。那么，押韵(有语音干扰)不押韵(没有语音干

扰)对判断成句不成句,应该是毫无影响的。如果从字到意义是经过语音这一阶段的,那么,语音的干扰就会影响到判断的速度。下面就是根据实验统计出来的结果:

	成　句	不成句	平　均
押　韵	2.555	2.395	2.475
不押韵	2.129	2.113	2.121
平　均	2.342	2.254	——

从表上可以看出,成句押韵的句子,反应时间需要 2.5 秒;成句不押韵的句子,反应就比较快,只需要 2.1 秒。总起来看,当有语音干扰的时候,反应时间就要增加五分之一左右。不成句的情况也一样:有语音干扰,反应时间就长;没有语音干扰,反应时间就短。

上面这两个实验否定了这样的说法:看汉字和看图画一样,是不通过语音的。实验结果证明,看汉字的时候并不能由字直接达到意义,而也是要通过语音的。

以上的实验引起很多人的兴趣。有些人就想进一步知道,汉字和英文虽然都和语音有关系,但是汉字终究离语音远一些,这种差别是不是能用实验来证实呢?宾夕法尼亚大学和天普 (Temple)大学的语言学家和心理学家曾经在一起对这个问题做了一个实验。①他们把接受实验的人分成两组,一组是中国人,一组是美国人,分别在香港和美国进行实验。用的方法和曾志朗的实验原则上一样,就是看语音干扰对判断句子的影响。比如,分别用幻灯片映出下面两句话:

(1)　A pier is a fruit。(码头是一种水果)

(2)　A banana is a fruit。(香蕉是一种水果)

---

① Rebecca A. Treiman, Jonathan Baron, and Kenneth Luk(陆仲真): «Speech recoding in silent reading: A Comparison of Chinese and English», (文字、语音与阅读理解:中英文的比较),«中国语言学报»1981,116 页。

(1)不成话，应当作出否定的反应；(2)成话，应当作出肯定的反应。但是，在英语里还有一个和pier声音近似的词pear(梨)。所以通常对于象(1)这类句子作出反应的时间就比较慢，因为在他的脑子里有另外一个声音相同或相近的词在干扰："A pear is a fruit"(梨是一种水果）就是成话的句子。他们在香港对中国人所进行的实验和以上完全是平行的。例如：

(1) 煤是一种水果。

(2) 铅是一种水果。

对(1)作出反应的时间比(2)也要慢。因为第一句话里有一个同音词"梅"在进行语音干扰。把美国人的一组和中国人的一组实验结果加以比较、统计，证实了这样的猜测：在阅读的时候，汉字和英文都和语音有关系，但是英文和语音的关系比较密切，所以在上面这种实验中，语音干扰对作出判断的影响要比汉字严重得多。宾夕法尼亚大学的这个实验很有意思。前些天我刚收到他们这篇文章，恐怕要到一年多以后才能正式发表。

下面简单介绍一下第二个问题，汉字和大脑的关系。

过去有不少人认为，象英语这样的拼音文字和左大脑的关系密切，是因为语音是由左大脑管的；汉字和右大脑关系密切，是因为汉字和语音没有关系，它和图画倒很接近，图画一般正是由右大脑管的。这种说法到底有没有根据？曾志朗最近也作了一个实验，① 叫作"两眼不同视觉实验" (dichoptic experiment)。

做这个实验需要一种很基本的仪器叫做 tachistoscope（简称T-scope)。我们知道，在看东西的时候，人们总是使视点(所要看的东西)、晶状体以及视网膜里的中央凹处在一条直线中，这样看得

---

① O. J. L. Tzeng(曾志朗)，D. Hung(洪兰)，B. Cotton and W. S-Y. Wang(王士元)，1979，«Visual Lateralization of China»，ed. by W. P. Lehmann, *Univ.* Texas Press, «Language» 56.197—202。

最清楚，因为中央凹在视网膜里是神经最多、最敏感的部分。两眼所看到的范围当然比视点大得多。以视点为中心，视区可以分成两半：左视区 (left visual field) 和右视区(right visual field)。在看视点的时候，视区内其他东西也是反映到两眼的视网膜里的，只是位置不同。比如左视区内出现一个“x”，它反映在右眼视网膜中央凹的右侧，视神经在这块视网膜上是透过去的。对左眼来说，也反映在它的视网膜中央凹的右侧，但是视神经在这块视网膜上是透不过去的。同样，右视区内出现一个东西，比如一个“o”，它也分别反映在两眼视网膜的左侧。但是，视神经在左眼这块视网膜上是透过去的，在右眼这块视网膜上是透不过去的。上述情况大致如下图。

由于神经构造的不同，左视区里的东西，先送到右脑，然后才送到左脑。相反，右视区的东西，先送到左脑，然后才送到右脑。许多人就是根据这个道理，研究相片是左大脑还是右大脑分析的，文字又是在哪一部分分析的，图画又是在哪一部分分析的。做这种实验，可以在同一个时间内，在左右两个视区分别出现不同的东西，然后问他看到了什么，看他反应的速度如何，看他看得对不对。这样，就可以测量出究竟是左大脑灵敏，还是右大脑灵敏。有人用这个办法进行实验，认为汉字象图画，放在左视区，到右大脑的

效率比较高；放在右视区，到左大脑的效率就略差。

曾志朗认为，对于语言来说，这种说法未免过于简单，何况汉字并不就是图画。于是他就作了一系列实验。实验分三步。第一步，用单字作材料，其中有的是象形字，有的是形声字。形声字里包含着声符，所以和语音的关系比较接近。然后用 T-scope 仪器进行两眼不同视觉的实验，得到的结果的确和上面那种说法一样，右大脑比较灵敏。但是，现代汉语里单音词毕竟是比较少的，绝大多数的词都是双音词。于是他又做了第二步实验，用双音词作材料，例如"黑板、桌子、粉笔、爸爸、妈妈、地板、风扇"等等。实验的时候，同时在左右两个视区放不同的双音词，让看的人立刻说出看到的词是什么。实验的结果非常有趣。和看单字完全相反。看单字的时候，右大脑灵敏；看双音词的时候，左大脑灵敏。也许有人说，因为要求看到以后立刻说出来，反应的速度受到左大脑语音部分的影响，所以右大脑不如左大脑灵敏了。为了反驳这种看法，就要作第三步实验，给左右视区同时放不同的字组，有的成词，例如"牛奶"，有的不成词，例如"铅奶"。但是，这次要求不做口头回答，而只是用手按电钮。如果认为成词，就用右手按，表示肯定；如果认为不成词，就用左手按，表示否定。有时又倒过来，用右手按，表示否定，用左手按，表示肯定。总之，要考虑得很周到，控制得很严密，否则得出来的结果就没有什么意义了。第三步的实验结果表明，确实还是左脑灵敏。

曾志朗从实验中得出的结论是这样的：过去认为图画和右大脑关系密切，语音和左大脑关系密切，实际上反映的是右大脑和左大脑在功能上的区别。左大脑一般讲来是比较有分析力的，右大脑一般讲来是接受整个印象的。所以，如果收进来的信号是一个整体（例如图画），不需要再进一步作深入的分析，右大脑起的作用就比较大；如果收进来的信号是一串东西，在时间上有差别，进入

大脑以后要把它的次序搞清楚，要进一步分析，那么左大脑的作用就是主要的。过去有人所以认为看汉字右大脑灵敏，并不是因为汉字和图画一样，而是因为他们的实验只用单字作材料，这种材料不需要大脑进一步分析，所以进来的这些信号主要在右大脑里处理。如果把字组成双音词作为实验材料，那么进入大脑的信号就有了时间上的差别，字和字之间的关系也需要进一步分析，所以主要就由左大脑处理。总之，信号的性质不同，在大脑里处理的部位也不同。我想，曾志朗的结论比简单地说学汉字是右脑灵敏还是左脑灵敏要深入得多、科学得多了。文字是这么复杂的一个系统，和语音又有密切的关系，因此不可能只由脑子里某一个很小的部分专管这类文字或那类文字，而是要由整个脑子协同处理的。但是，由于不同信号的不同性质，大脑各部分功能的比重是有主次之分的。

关于汉字和大脑不同部位关系的问题，当然还有待于广泛深入地进行研究。曾志朗的说法，我现在还很难说是不是完全同意。但是我觉得这个实验非常有意义。值得再进一步做很多这样的实验。有了曾志朗这个实验，我们至少不会象过去那样把问题简单化了。我觉得，研究汉字是一项迫切需要进行的工作。汉字不但影响那么多的人，而且对语言理论的研究也很有意义。因为自从有了文字以后，语言就有了耳朵和眼睛两个渠道。从两个角度——耳朵和眼睛——去研究大脑是怎样处理语言信号的，收获一定比只从一个角度研究要大。对我们来说，更重要的是这方面的研究可以使我们对汉字有更透彻、更深刻的了解，从而为国内的文字改革方针打下坚实的科学基础。这是中国语言学对中国的社会可以做出贡献的一个地方。

四　近几十年来的美国语言学

语言学近几十年来的发展，我以为主流是在美国。日本、英国、法国、瑞典等国，对语言学，特别是对语音学、心理学等等方面，也都有他们的贡献。但是投入人力最多，工作最活跃，文章发表得最多的，仍然要推美国。

近几十年来美国语言学的发展，可以分为三个时期。学术的发展总是逐渐进行的，划分时期不能象回忆孩子的生日那样有一个准日子。但是如果一定要在近几十年美国语言学发展史中找出一个重要日期，那就可以说是 1933 年。那一年 L. Bloomfield 出版了他的《语言论》(Language)①。

二十世纪前半叶，在美国语言学界有两位大师，就是L.Bloomfield 和 E. Sapir。他们两人学识渊博，影响巨大，但是作风很不相同。Bloomfield 不大和人接触，很少外出，也不常和自己学生一起工作。他对语言学的影响大部分是通过他的著作，特别是 1933 年出版的《语言论》那本书，任何对语言学有真正兴趣的人都是非看不可的。Sapir 也写过很多著作，他在 1921 年出版的《语言论》(Language)② 虽然只有一百多页，但是高瞻远瞩，写得非常之好，到现在还经常被人用来作为语言学导论的教科书。他很有人缘，美国语言学家中有很大一部分人或者是他的学生，或者是他学生的学生，或者曾经和他一起做过研究工作。他的学生比较著名的有 M. R. Haas, C. F. Voegelin, M. Swadesh, B. Bloch, Z. S. Harris, C. F. Hockett 和李方桂等人。

---

① 汉译本由袁家骅、赵世开、甘世福译，商务印书馆出版，1980 年。

② 汉译本由陆卓元译，商务印书馆出版，1964 年。

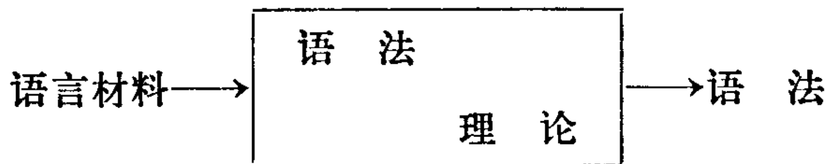
起初，美国语言学界人很少，活动也不多。1924 年底美国语言学会 (Linguistic Society of America) 成立时，会员才不过二百多人，开会往往只有几十人参加。美国各大学当时还没有独立的语言学系，一些语言学家都是在别的系工作的，比如 Sapir 是在人类学系，Bloomfield 是在德语系。甚至语言学本身在很多人的头脑里也是不存在的。后来，有很多外来因素使语言学的活动变得活跃起来了。

最主要的外来因素是三、四十年代的第二次世界大战。美国本来和外界接触不太多，但战争一爆发，要派军队到太平洋上的许多岛上去，到非洲和南美去，那就不仅需要了解当地的社会和文化情况，还需要了解当地的语言，甚至需要教军队学说这种语言。为此，国防部和国务院都拨出了大量经费，开办了许多语言学校。这就给语言学家带来了很多工作，比如 Sapir 的学生 Bloch 研究日语，Hockett 和 Swadesh 研究汉语，等等。语言学家因此学到了很多一般人接触不到的语言，这对语言学工作无疑是一个很大的促进。

还有一个外来因素是传教。这在一般人是不大能想象的。美国现在的语言学组织中，除了美国语言学会有好几千会员以外，最大的就数传教组织了。其中最重要的是语言学暑期讲习所 (Summer Institute of Linguistics)。它和美国语言学会每年夏天在各大学轮流举办的语言学讲习所 (Linguistic Institute) 完全不是一回事。语言学暑期讲习所是传教组织，和《圣经》翻译组织(Wycliffe Bible Translation) 都是教会办的，目的是训练到世界各地去传教的人分析当地的语言，并且把《圣经》翻译成当地的语言。这个讲习所规模很大，每年夏天往往要在世界各地分十几处举办。著名语言学家 K. L. Pike 曾长期担任这个讲习所的领导，《圣经》翻译则由 E. A. Nida 负责。这两个人都是在语言学上很有贡献

的，都培养了不少语言学家，也都写了不少著作，其中 Pike 的《音位学》(Phonemics) 和 Nida 的《形态学》(Morphology) 都是大家很熟悉的，也是学语言学的两本主要教科书。

语言学暑期讲习所的看法和国防部那些语言学校的语言学家有很多地方不同。美国语言学从三、四十年代开始逐渐分成了两派，一派有 Pike, Nida 这些人，另一派有 G. L. Trager, B. Bloch, C. F. Hockett, M. Joos 这些人。四十年代有两本最重要的书，正好能够代表这两个学派。一本是前面已经提到的 Pike 的《音位学》，另一本是 Bloch 和 Trager 的《语言分析纲要》(Outline of Linguistic Analysis)①。我们现在管那个时代的语言学叫"描写语言学"(Descriptive linguistics) 或"结构语言学"(Structural linguistics)，其实二者并不是完全一致的。"描写"(descriptive) 和"结构"(structural) 这两个形容词正好有两本很重要的教科书对它们加以阐发。一本是 H. A. Gleason 的《描写语言学导论》(Introduction to Descriptive Linguistics)，Gleason 也是一个传教士，和语言学暑期讲习所等组织关系很密切。另一本是 Z. Harris 的《结构语言学的方法》(Methods in Structural Linguistics)。在 1933—1957 年间，描写语言学和结构语言学这两个名称有时候是可以通用的。当时在这些人看来，语言学就是一套方法。比如，Pike 那本书的全名是《音位学——把语言写成文字的技术》(Phonemics, A Technique for Reducing Languages to Writing)。不只是 Pike，当时许多人都是这种看法。他们有时候画出这样的图来说明他们的看法：



① 汉译本由赵世开译，商务印书馆出版，1965 年。

他们认为，语言理论就是一套方法，这套方法如果研究得好，只要向它送进语言材料去，它就能象电子计算机的程序那样，一步一步告诉你怎么做，最后得出来的就是语法。因此有人就认为，从结构语言学的观点来看，语法理论只不过是从语言材料中发现语法的一套方法。在他们的著作里，“技术”(technique)，“程序”(procedure)，“方法”(method) 这类词用得很多。我想，他们之所以产生这种看法，是和他们的工作性质有关的。因为，不管是教军队学那些一般人很陌生的语言，还是把《圣经》译成这些语言，都需要到一个地方去和一个以前从来没有人分析过的语言接触，并且在最短期间——几年，甚至几个月——把它分析出来，然后尽快写出一部语法。他们工作的重点就是训练工作人员去分析语言材料，他们最迫切需要的就是一套方法。

他们在理论上也很有贡献。我们现在经常谈到的语言构造的单位，比如音位 (phoneme)，语素 (morpheme)，语素音位 (morphophoneme)，等等，差不多都是在那个时期才逐渐弄清楚的。象赵元任先生的那篇《论音位标音法的多能性》(The Non-uniqueness of Phonemic Solution of Phonetic Systems) 也是在结构语言学的环境里，在那样的条件下写成的。所以可以说，二十世纪的美国语言学，基础是由结构语言学学派奠定的。在这以后，我们虽然做了一些进一步的比较理论性的工作和比较高级的实验研究，但是没有这个基础还是做不成的。

不过也许可以说，他们最大的贡献还是在于为我们搜集了好几百种新的语言材料。在这以前，语言学研究的对象几乎总是局限于英语、德语、法语等少数几种常见的语言。而后来，由于翻译《圣经》的需要，由于军队的需要，许多人员到遥远的地方去，深入到森林、沙漠地区，了解到以前从来没有人分析过的成百上千种语言，从而使我们对语言的认识有了广阔的基础。正因为如此，我们现

在才有可能提出所谓“语言普遍性”(language universal) 的问题，这不是从过去的三五种语言中能找出来的。

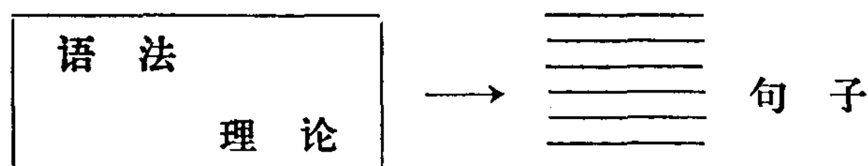
但是，按照所谓的“发现程序”(discovery Procedure) 来看，在分析比较复杂的语言现象的时候，他们的那套方法有时就会遇到困难。从那几十年的文献来看，他们比较注意分析语音、语素、词组屈折；而到了分析句法的时候，牵涉到词组和词组之间的关系的时候，就比较忽略了。没有人写出过一本内容丰富的分析句法的教科书，因为当时对句法应该怎样分析还不很清楚。尽管 B. Bloch 和 R. S. Wells 也写过几篇文章谈词组应该如何分析，如何图解，词和词之间应该如何划分界限，但是现在看来，那都是一些很小的进步。从整个语言系统看来，结构语言学派在句法上的贡献比较最小。

当时也确实有人眼光放得比较远。比如 Z. Harris, 他在研究音位学和词汇学的同时，对于句法也很有兴趣。Harris 的数学基础很好，他看到传统语法书总是分章叙述否定句、疑问句等等不同的句型，觉得这样分别叙述，很难注意到不同句型之间的关系。他认为，句子和句子之间的关系恰恰是整个语法系统中很重要的一个概念。五十年代他接连写了好几篇有关句法分析的文章，提出用转换 (transformation) 的办法把不同的句型联系起来。这是他的一个很大的贡献，然而在当时并没有引起人们的注意。

1957 年，Harris 的学生 N. Chomsky 出版了《句法结构》(Syntactic Structures) 一书，立刻轰动一时。这本书宣告了结构语言学从此没落，Chomsky 所提倡的语言学理论开始兴起。[①] Chomshy 继承了 Harris 句法研究方面的很多概念，并且加以发

① 该书有王士元、陆孝栋以《变换律语法理论》为名的编译本，香港大学出版社，1966 年。该译本列举了汉语例句，并有较详细的解释。最近国内有邢公畹的译本，中国社会科学出版社，1980 年。

展，提出了许多和结构语言学根本不同的看法。他认为，只把语法理论看成是一套方法是不够的，语言中有种种微妙的关系，用什么方法去发现它们，这并不重要，重要的在于所发现的东西是否正确。他的看法可以用下图来表示：



也就是说，语法理论不是一套方法，而是一套假设，它能使我们推断出来的东西是一个一个的句子。简而言之，语法理论就是产生句子的句法功能。

Chomsky 在语言学方面受 Harris 的影响很深，此外他的数学基础也非常好，所以他想把数学的基础概念用到语言学里来。谁都知道，数学总有一些不需要证明、也不可能证明的假设 (postulates)。有了这些假设，就可以逐步推导出原理 (theory) 来。比如说，一，二，三，四……这些数目看来好象很好理解，但要准确地、客观地说明它，却相当困难。因此数学家就先作出一系列假设，然后把数目一个一个地推算出来。Chomsky 认为，语法也是一套规则，就象数学里的一套假设一样。从这个语法里可以推导出一个一个的句子。如果做得对，只要是句子，就都能推导出来，而不成句的不合理的词串，就一个也不应当推导出来。在任何自然语言里，句子显而易见都是无限之多的。数学里没有最大的数目，因为任何一个数目都可以再加上另一个数目。同样，语言里也没有最长的句子，因为任何一个长句都可以再加上点什么。因此，有限的符号，可以反复运用，生成无限多的合法的句子。这就是所谓“生成语法”(Generative grammar) 的最基本的一个原理。前面已经谈到，Chomsky 还继承了 Harris 的转换理论，所以这一派的语法

就叫做“转换生成语法”(Transformation Generative Grammar)，有人把它简写成 TG。

我们可以举一个具体的例子来说明。比如英语可以说 We go, We can go, We went, We could go, 可是不能说 *We can went, *We go can。仅只这么几个词的组合，就是有一定的规则的。Chomsky 在语法分析中使用了一些符号，最基本的一个符号是 S，代表句子。在大多数情况下，句子可以分成两部分，一是主语，名词组 (NP)，一是谓语，动词组 (VP)。在前面举的那些句子里，我们把名词组译成 We。当然名词组也可以是很复杂的。动词组分成助动词 (Aux.) 和动词 (V) 两部分。以上可以写成：

S→NP+VP ①

NP→We ②

VP→Aux.+V ③

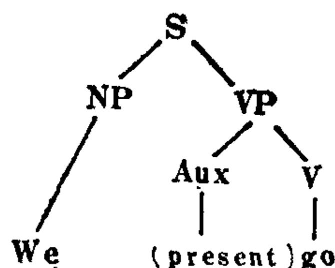
在生成语法理论里，也许最中心的一个概念就是语法规则。有种种不同的规则。而有了种种不同的规则，就有种种不同的功能。上面的规则叫做“短语结构规则” (Phrase-structure rule)，它把一部分一部分的词分成组合。用这个规则的时候，一开始左边总是只能有一个符号，右边可以有一个、两个、三个以至更多的符号。上面的规则 ① 表示句子由主语和谓语组成。规则 ② 表示主语是 We。规则③表示谓语由助动词和动词组成。助动词有四个部分，最前面的是时态(tense)，然后是三个可能出现的部分：情态动词 (modal)，完成体 (perfective)，进行体 (progressive)。这三部分的次序是非常重要的。这条规则可以写成：

Aux.—T+(modal)+(perf.)+(prog.) ④

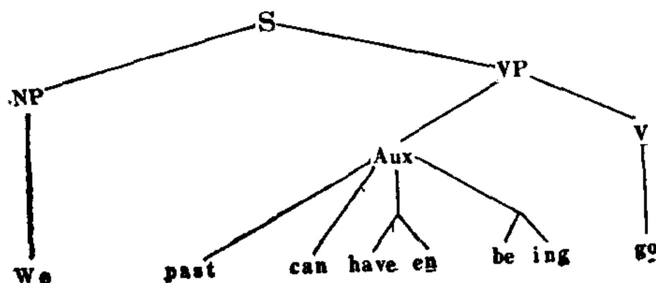
括号是表示可有可无的。我们可以算一算究竟有多少可能出现的情况。时态一共只有两个：过去 (past) 和现在 (present)。情态动词有好几个：will, shall, can, may 等等，我们姑且只算五个。

最后两项都只有有或者没有两种可能。那么，要产生助动词这个短语，至少是 2×5×2×2=40,有四十个不同的可能性。

如果我们选的动词是 go，助动词是现在时，则结果最简单，得出来的句子就是 We go。用树形图表示如左下。



如果助动词选的是过去时，情态动词是 can，再加上英语完成体和进行体都是不连续结构 (discontinuous construction) have-en 和 be-ing，情况就复杂多了。按照短语结构规则，应该产生出这样一串：We past can have en be ing go。用树形图表示如左下。



这自然不成句子，需要进一步用转换的办法来处理。转换和短语结构规则的基本不同是左边可以不只是一个符号。有一个专门处理助动词的转换规则，叫做 T Aux.，它的内容是：如果一个词缀 (affix) 后头有一个动词，就要把这个词缀移到这个动词后头，然后用符号#把它堵住，使它不能再向后移。这意思可以表示如下：

Affix－Verb＝>V　Affix #

在上面的例子里，Past、en、ing 是词缀，（从词位观点看，past 也是一个词缀）can、have、be、go 是动词，也就是：

	A	V	V	A	V	A	V
We	past	can	have	en	be	ing	go

第一次用 T Aux. 转换规则，结果是：

V A V A V A V

We can past # have en be ing go

这个例子里还有两处应该用 T Aux. 转换规则，运用以后，结果就是:

V A V V A V A

We can past # have be en # go ing #

最后就成为:

We could have been going

由此可见，小小的一条转换规则能起很大的作用。它能产生很多不同的句子，可是不会产生不合语法的句子，如象* We have could been going, *We could have going, *We could have go 之类。这种转换规则还有一些别的作用。例如，英语的疑问句一般总是把时态提到前面，总是 Does he go, 而不是 *goes he。为什么要用 does 呢？因为时态提前以后，没有动词支持它，于是就加了一个 do, do 是在任何情况下都有同样的搭配能力的。

Chomsky 在 1957 年写的这本小书的确有它精彩的地方。他使句法理论有了一个新的系统。但问题也慢慢产生了。一开始人们还弄不清这些问题是原则上的问题还是理论不够完善的问题。Chomsky 自己的办法是修补。他每写一篇文章，就等于是给他的理论打几个补丁。但结果是顾到这个，顾不到那个，形成一种捉襟见肘的局面，语言学理论也由此形成了一个越来越乱的局面。虽然有人在开始时曾说这本书是语言学理论上的一个"革命"，但也有人（如语言学家 C. F. Hockett）说这本书对语法理论起了"毁灭"的作用。

Chomsky 的语法理论经过了好几次根本性的改变。例如，1957 年他在《句法结构》里提出的只有两种基本上不同的转换：一种是简单转换 (simple transformation)，左边只能有一个符号；另

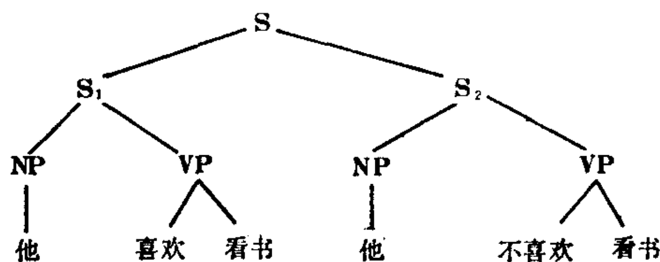
一种是一般转换 (generalized transformation) 或变异转换 (modification transformation), 左边也可以有好几个符号。到 1965 年发表《句法理论问题》(Aspects of the Theory of Syntax) 的时候，他在短语结构规则里又加进了一些新东西，这时他认为产生无限多的句子主要是两种情况: 连接的 (conjoining) 和嵌入的 (embedding)。前者如"张三弹琴，李四喝酒"; 后者如"我买了一本张先生写的书"。连接的情况可以表示为 S→s+s。我们可以用汉语的例子来说明。汉语有这样一种疑问句:

他去不去?

他看不看电影?

他喜欢不喜欢看书?

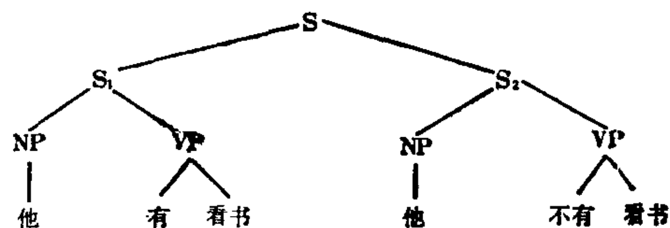
我们以第三句为例分析一下。这个句子可以表示如左下。



在传统语法里，"他喜欢看书"和否定句"他不喜欢看书"、疑问句"他喜欢不喜欢看书"是分开叙述的。在转换生成语法里，这三个句子则是由一个句原演变来的，应该表达出它们之间的关系。在这里要使用短语结构规则，然后再使用转换。但是，现在所需要的已经不是刚才所说的移动词组的转换，而是一种省略词组的转换。省略是语法里一个很重要的程序，语言之所以能够这样丰富，其中一个原因就是我们能够把几个不同的概念"挤"在一个短的句子里。所谓"挤"，就是把多余的省略掉。比如前面提到的"我买了一本张先生写的书"，就是把"我买了一本书"和"张先生写了一本书"两个句子挤在一起省略掉一个"一本书"的结果。上面树形图里的句子，首先可以省略的是 S_2 里重复的 NP "他"，成为"他喜欢看书不喜欢看书"。其次，两个

"喜欢"的宾语可以任意省略一个：如果省略前一个，就成为"他喜欢不喜欢看书"；省略后一个，就成为"他喜欢看书不喜欢"。这样，各种句子之间的关系就可以连起来了。

如果我们改变一下这个句子的情况，还是可以用类似的办法来处理的。比如，把现在时态改为过去时态，那么，在某个生成的阶段，原来"喜欢"的位置上就应该出现一个表示时态的词，原来的树形图就变成如右图。这样生成的句子至少在普通话是不说的。但为了保持和刚才分析的那些句子的共同点，我们还是把它分析成上面树形图那样的深层结构 (deep structure)，然后再增加一些规则进去。比如，如果"有"的前面没有别的东西，就把它移到动词后面，变成一个"了"，"有看书"就成为"看了书"；如果有"不"在"有"的前面，"不"在很多情况下会变成"没"，这是语素音位叙述的一条规则。于是全句变成：



他看了书，他没有看书？

再进一步省略 S_2 里的 NP "他" 和 VP 里的宾语 "看书"，最后就产生了一个很常用的句子：

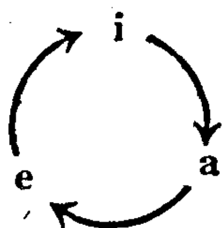
他看了书没有？

这种省略是在历史上一步一步变化而来的。有的方言在转换以后得到的是"他有没有看书"这样的句子。我们不能说别的方言一定也经过这个阶段，但至少可以说是有这种可能的。这个问题我在另两篇文章里也谈到过。①

---

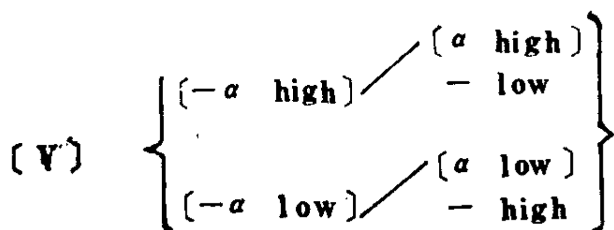
① William S-Y. Wang(王士元)，«Two aspect markers in Mandarin»(普通话两种体的标志)，«Language» 41.457—70，1965，«Conjoining and deletion in Mandarin Syntax» (普通话句法中的连接和省略)，«Monumenta Serica»26.224—36，1967。

转换生成的理论在语音方面也有很多贡献。实际上，近年来这个学派的著作看的人最多的也许是一本语音方面的书，那就是 1968 年 Chomsky 和 M. Halle 合写的《英语的音型》(The Sound Pattern of English)。我们可以举一个语音方面的简单的例子来说明。英语由中古到现代有种种音变，其中有一种很有名的音变，Jespersen 管它叫做"主要元音转移" (the great vowel shift)。这是指一种音变在词汇里不是同时完成，而是逐步在变的，现在我们在某些词里还可以看到这种音变的不同阶段。以前元音为例，"主要元音转移"的变化是 [a] 变 [e], [e] 变 [i], 然后有趣的是 [i] 又变回来成为 [a]。例如 sanity 的元音是 [a]，相应的形容词 sane 是 [e]; clean 的元音是 [i]，相应的动词 cleanse 是 [e]; five 的主要元音是 [a], 后面加上 -th 变成 fifth, 元音是 [i]。这种随着环境变化而改变元音的词有好几百个，变化很有规律，可以用左下图来表示。生成语音学在分析这种语音现象时，先

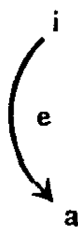


把这三个元音分成高 (high) 和低 (low) 这两个不同的区别性特征，然后用一套有次序的规则加以处理。

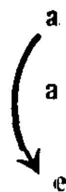
我们就以 five 和 fifth 作为例子。从语素音位学 (Morpho-phonemics) 的观点来看，它们应当是 fīv 和 fīv+θ。按照 Jespersen 所发现的"元音缩短" (Vowel shortening) 规则，长元音如果后面有两个阻塞音或两个不同音节，就要缩短。因此 fīv 的元音不变，fīv+θ 的元音要缩短变成 fivθ。然后 fīv 再根据"音渡插入" (Glide insertion) 规则加进一个音渡。所加音渡的舌位前后应该和元音舌位相配。所以 e+i→ei (不可能是 eu), o+u→ou (不可能是 oi)。fīv 加上音渡就成为 fījv。最后，再经过"元音转移" (Vowel shift) 规则。这一规则分成两段，可以用下页左上表来表示: 第一段是 −low 的元音，把它的 −high



的那个区别性特征改变：如果是＋，就改成－，如果是－，就改成＋。第二段是－high 的元音，把它的－low的那个区别性特征改变：＋改成－，－改成＋。根据规则可以看出如左下图所示的音变过程：第一步，[i]变成[e]，[e]变成[i]，但是不影响[a]；第二步，[a]变成[e]，[e]变成[a]，[i]不受影响，因为它是－low。因此，fījv 里的[i]变成[a]，整个词就变成了 fajv。这样，five 和 fifth 之间的关系就联系起来了。







不过，对语音这样分析也有很多弱点。例如，区别性特征只分成高和低两个，这样最多只能区别元音的三个不同高度。（虽然在逻辑上有四种可能，但是兼有 ＋high、＋low 这两个区别性特征的元音事实上是不存在的。）由于篇幅关系不能在此详述，请看我在另一篇文章里提出的见解①。实际上很多语言里元音高度的区别不止三个，而可能有四个甚至五个。这就是一个很基本的弱点。此外，还有人指出，《英语的音型》这本书还有不少逻辑性的矛盾。

从 1957 年到 1968 年《英语的音型》出版后这段时间，可以算是转换生成语言学派比较兴盛的时期。但是由于它存在着原则性的问题，弱点自然会逐渐暴露出来。补救是无济于事的，因为补一处，破十处，弱点反而越来越明显。在《英语的音型》出版后没有几年，出现了一本名叫《〈英语的音型〉评论集》(Essays on the Sound

---

① William S-Y. Wang（王士元），«Vowel features, Paired Variable, and the English Vowel Shift»（元音特征，成对变项和英语元音转移），«Language», 44, 695—708, 1968。

Pattern of English) 的书，其中包括不同的人写的几十篇文章。这些文章指出了《英语的音型》中的种种弱点，并认为这个系统不值得去补救。应该承认，转换生成理论在句法和语音方面是有贡献的。但是必须指出：第一，这些贡献很多是继承了前人的成果，并不是他们首创的。第二，他们在理论上一个很基本的假定，就是能用一套完整的规则生成无限多的句子，这个目标是定得太高了。就我们现在对语言的认识来看，这是不可能达到的。第三，也许是最基本的一个弱点，就是用这样的观点去研究语言，必然会一步步走向抽象，和语言本身以至社会现实越来越脱节。比如有一些在麻省理工学院念书的学生，除了英语之外，连声学也不怎么学，成天关在屋子里画箭头，钻那些公式的牛角尖，结果当然钻不出什么东西来。又比如我自己，1957 年看到 Chomsky 的《句法结构》时，由于年轻，容易冲动，觉得这本书已经把语言学里的所有问题都解决了，于是用了许多年的功夫钻研它，并且把它译成了中文。但是，越是深入一步，就越是发现问题并不那么简单。语言是那么复杂的东西，跟人的生理和社会交际都有密切的关系。只抱着一本书在书房里，老是那么写公式，老是那么钻，是没有太大前途的。语言本来不是一个代数系统，它是活生生的东西，是经过几十万年演变而来的、非常奥妙的活生生的东西。语言比代数系统、逻辑系统要复杂得多。转换生成理论看不到这点，这就钻进了一个钻不出来的牛角尖。大约从六十年代后期开始，很多人对它感到失望，逐渐离开了它。

从 1957 年到 1973 年是美国语言学发展的第二个阶段。1973 年是一个很重要的转折点，那年夏天美国语言学会在密执根 (Michigan) 大学举办了一年一度的语言学讲习所。这一期语言学讲习所和前几年的情况大不相同。前几年讨论的总是那些老东西，哪个区别性特征前面加“+”，哪个前面加“−”等等，总是钻牛角

尖。这一次相反,讨论得非常生动,明确提出了语言学中存在的中心问题。过去那些过于抽象化的语言学派把历史语言学、方言、社会语言学等等都给排挤掉了,这次明确提出,研究语言的时候,不能把它当作一个抽象的系统来处理,而一定要注意到语言在时间上和地域上的变化,要注意到它和社会的关系。

从那时起,美国的语言学界变得活跃多了,不象过去只写抽象的公式时那样单调了。思想开始活跃起来,各种人有各种不同的想法。在这个阶段,应该提到 William Labov, 他在语言学的方法和理论、方言、语音、语义、特别是英语的音变方面,都有重要的贡献。1973 年的语言学讲习所就是根据他的意见办起来的。

Labov 有一个很重要的贡献,就是他提出了语言的变项规则(Variable rule)。他认为语言这个活生生的东西并不是转换生成理论那些死板的规则所能表达的。不同的人之间,语言就有所不同;同一个人在不同的情况下,比如在家和孩子说话,出门做学术报告,所用的语言也有很大不同。Labov 认为,根据语言的不同的变化,首先可以把它写成一种统计性的变项规则,然后可以根据这种规则去判断语言变化的趋向。比如,英语的词尾 [-d] 在很多情况下被省略: d# → ∅,但是这个省略情况相当复杂。例如 find Ann 和 find Ben, 前者后面是元音,后者后面是辅音,情况就不同;如果前头有个语素的界限,如fined Ann 和 fined Ben, (fined 是 fine 的过去式)情况又不同。同样一个 d#, 有时是在元音前,有时是在辅音前,有时是词的一部分,有时又是单独一个语素,但在这些不同情况下,它的省略与否仍然是有规则可循的。根据他的统计,同是一个 d#,在辅音 (C) 前比在元音(V)前省略的频率高;而如果它是一个语素,因为含有语义,省略的频率就要低一些。因此,写这种规则的时候,不能只是简单地说“省略”或“不省略”,而应该把种种不同环境都写进规则去,还应该加一个统

计方面的引证。Labov 还发现，这种变化也是随着不同的人和不同的语言环境而变化的。青年人省略得多一些，年纪大的省略得少一些，男人和女人，不同文化程度的人，情况也都有所不同。这样，他就逐步把语言和社会上的各种因素联系起来了。这是很重要的发展。Labov 是主张在社会环境中研究语言的。他认为，研究语言的人不能把自己关在书房里幻想，一定要到外面去，一定要注意人们在使用着的、活泼的语言。他自己就总是随身带着录音机，随时录下音，带回去分析。

从 1973 年开始，随着上述新观点的提出，美国语言学可以说是进入了第三个阶段。也出现了一些新的刊物，如《大脑和语言》(Brain and Language)，《语言和社会、社会语言学》(Language and Society, Sociolinguistics) 等，支持上述新观点。语言学家也逐渐和社会上一些别的单位和组织建立了更密切的联系。我觉得，这是一种健康的、可喜的新发展。在此以前，语言学界和其他部门，甚至和语言教学方面的人，都是界限分明，互不相关，似乎你是你，我是我，你搞你的教学，我搞我的抽象理论。这种很糟的情况，现在已经有了很大改善了。过去有些学问曾经被认为是“低级”的，比如应用语言学 (Applied linguistics)，在二十年前就被人看不起，似乎一个人如果学问搞得好，就不应该去搞它，而应该去搞抽象语言学或语言学理论，可是现在又有很多人注意它了。语言学家不只注意语言教学、儿童学话等等，还开始到医院去和医生合作，看看在语言障碍、失语症、阅读困难等等病理现象方面，语言学能够提供些什么帮助。这样，语言学的基础就越来越广阔了。

在结束我对近几十年来美国语言学情况的介绍之前，我想谈一些不太成熟的想法，可能表达得不够准确，但是我想大家能理解我的意思是诚恳的。

首先，我觉得目前中国语言学界最好能把工作中心放到具体

问题的研究上，不要去走别人走过的弯路，那种钻牛角尖的、太抽象的道路。我们应该从那里吸取教训。研究具体问题，总要比搞抽象的、全盘性的理论好得多。比如汉语方言里的声调，量词的系统和它在方言里的分布，以及各民族语言间因相互接触而产生的语言变化，等等，就都是很好的研究课题。我们研究行为科学、社会科学的人，有时应该和一些比较更为成功的学科，比如数学、物理学比较一下。他们比我们成功自然有很多原因，其中一个很重要的原因，就是他们研究的对象比我们的简单得多，可以控制的因素多得多。研究一种东西，再没有象研究人那么复杂的。但是，还有一个方法上的原因值得我们注意，那就是他们总是研究具体的问题。象前些年语言学在美国那样，每隔几年就出来一个人宣布：我有一个理论可以解决语言学上的所有问题，这种情况在比较成熟的学科里是没有的。实际上，美国语言学在五、六十年代的那种情况，只有缺乏经验的年轻人听了才会高兴。但往往不过几年，理论的漏洞就暴露了出来，从长远来看，贡献就不那么大了。

我经常和一位研究数学的好朋友讨论学术问题，但有时总象是谈不到一起。有一天他对我说，他知道语言学和数学的根本不同是在什么地方了：数学的发展至少已经有两千多年的历史，每一代都是踏在前辈的肩上再前进一步的，语言学家则好象不是这样，有时候似乎是喜欢踩在前辈的脸上。这话是有些道理的。我以为，我们不能把做学问弄成做买卖那样，不应当有商业上的竞争性，一定要共同合作，互相帮助。真正做学问的态度不应当是看谁的东西风行一时，而应当是看谁能够留下一点哪怕很小的东西作为长远的贡献。我想这样一种合作的精神，在社会主义国家应当是特别容易得到理解和发扬的。

为了能够做出真正的贡献，选择研究课题就非常重要。研究问题需要时间。但是，有时由于性急，可能只把两成时间花在选题上，

八成时间花在研究这个问题上，结果往往会在完成以后才发现那个问题并不是很有意义的。所以，我想宁可在打基础的时候走得慢一些，把问题选好，目标看准。这样，即使钻研出来的是个小问题，但是小问题有时候也会有很大的意义。比如，看起来也许仅仅是是在研究汉语方言的塞音，但是很可能因此能对语音理论作出很基本的贡献。不过，也不应该是哪一位名家在搞什么，我们就也去搞什么。我们有自己的环境，自己的材料，自己的传统，应当研究我们能做出特殊贡献的问题。我想到一个笑话。有一个人在意大利旅行的时候买了一条裤子，样子特别好，他很喜欢。几年后穿旧了，他就找裁缝替他比着旧的一模一样地做一条。几天后他去取裤子，打开一看，吃了一惊，原来裁缝按照旧裤子的样子连补丁也给做上去了。所以，别人走错的路，选错的问题，我们都应当看清楚，不要再花时间和精力去打那种"补丁"。

最后，人们近几年逐渐了解到，语言这个东西实在太复杂了，单从语言学这个狭窄的角度来研究，收获总是有限的。所以，现在已经很少还有语言学家单独自己进行研究的了，大多数语言学家都在设法和别的学科取得联系。把语言学和别的学科之间的关系打通，看来是对语言学家的一个基本要求。考古学里有语言学，哲学里有语言学，物理学、心理学等等也都是如此。因为谁都要用语言，所以很多学科都对语言有兴趣，同时也可以对语言学做出贡献。就象建筑房屋一样，地基挖得越宽、越深，楼房就能盖得越大、越高。我们当前的责任，就在于把语言学和别的学科联系在一起，把它打通。如果打通了，我们的基础就会变得更加广阔，我们语言学的前途就会更加光明了。

作者后记

一九七九年夏，我有幸回祖国到北京大学作了为期两个月的学术报告，受到北大中文系师生的热情接待，情景至今如在眼前。

我在北大作了关于实验语音学和近年来语言学发展概况等两个方面内容的学术报告。北大中文系语音实验室把我的报告录了音，在我返美后不久就寄来了记录稿，并告诉我准备出版。北大的朋友们如此花费精力热情地帮助，使我十分感激。我本应尽早把这个记录稿从头至尾仔细地加以整理，以报答他们对我的期望。但因琐事缠身，一搁两年，一直未能抽出时间来实现这一愿望。现在不能再拖下去了。我谨抱着抛砖引玉的希望，把这个稿子稍加整理，让它出版。

由于我在中学时代就离开了祖国，现在用中文来表达学术上的观点是有一定困难的。幸亏这个记录稿由北大中文系语音实验室做了仔细的修改，实验室负责人林焘先生和王福堂、王理嘉两位先生为此付出了辛勤的劳动。我在整理时，又承张连生女士协助，补充了一些插图及表格。我谨对他们表示深切的谢意。但我想一定还有很多错误及不妥之处，希望读者指正。

中国的语言学是有光辉的传统的，尤其是对语音演变的基本规律认识得很早，在这方面留下了许多宝贵的文献。活的语言材料在中国更是丰富，汉语众多的方言，再加上数十种少数民族语言，都为我们中国语言学工作者提供了广阔的天地。我衷心地盼望祖国从事语言学研究的朋友们能够尽快地探索和使用最新最有效的方法研究这些材料，使中国的语言学在对语言的了解方面做出应有的贡献，进入世界语言学研究的前列。

王士元　1981年9月